Scalable and Interpretable Graph Modeling with Graph Grammars

---

Satyaki Sikdar

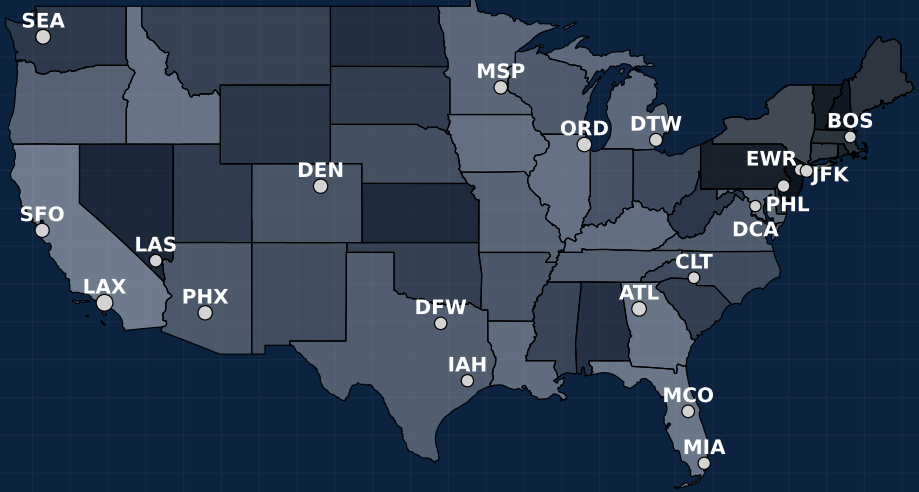
Ph.D. Defense

Committee: Dr. Tim Weninger, Chair

Dr. David Chiang · Dr. Peter Kogge · Dr. Danai Koutra

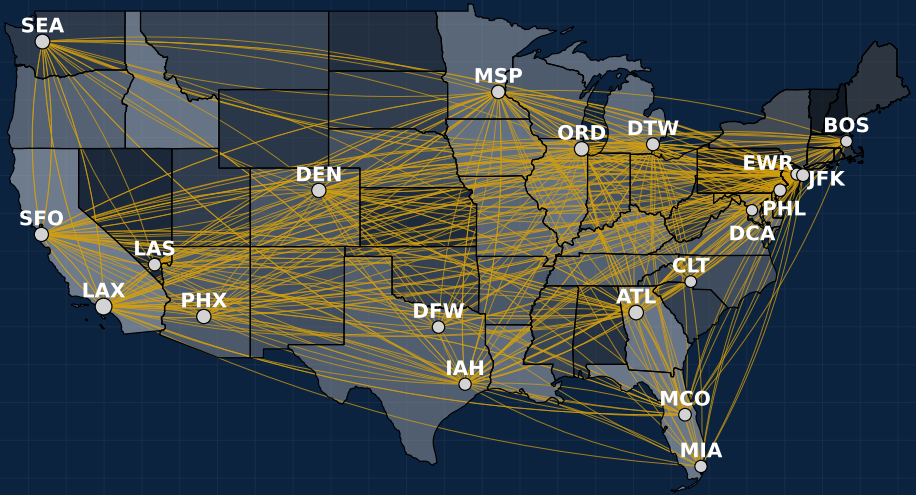
Introduction

US Flights Network in 2011

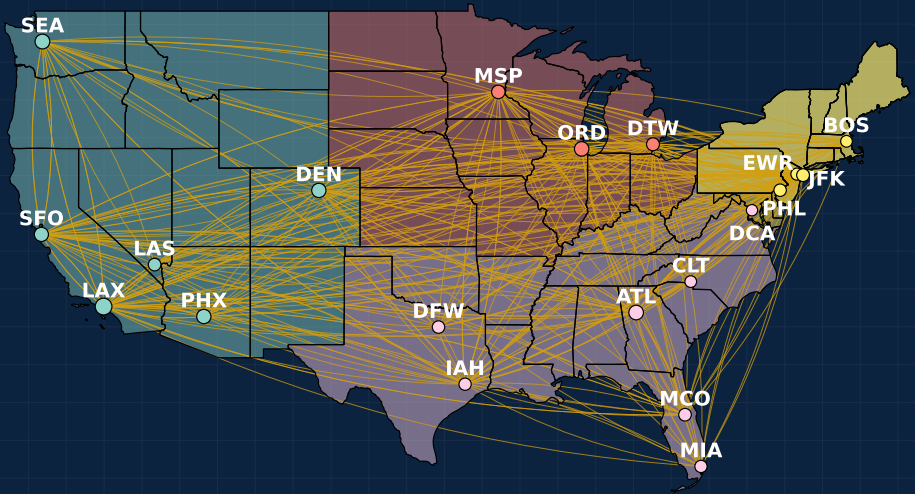


Introduction
Definitions
Graph Models
Overview
String Grammars
Graph Grammars
Grammar Extraction
Graph Generation
Infinity Mirror Test
Wrapping Up

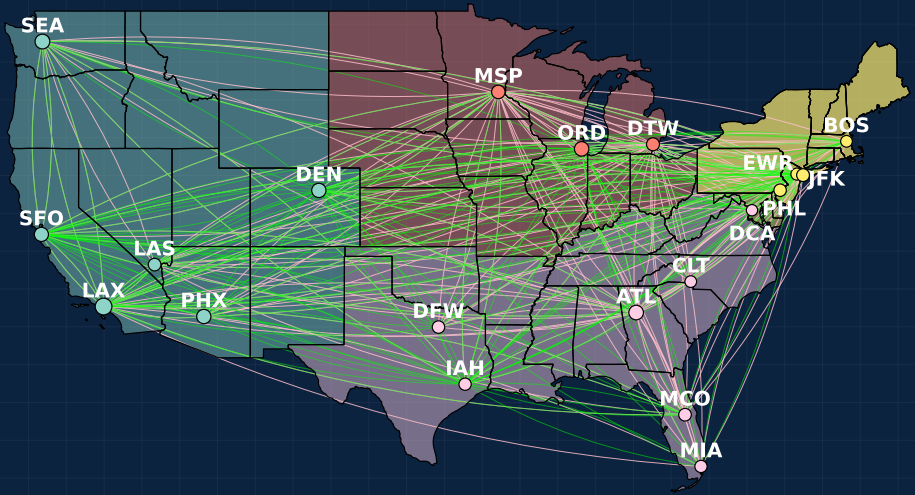
US Flights Network in 2011



US Flights Network in 2011



US Flights Network in 2011



How to Study Graphs?

Subgraph Mining

- Extract *statistically significant* patterns

How to Study Graphs?

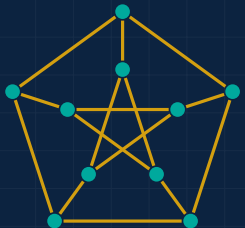
Subgraph Mining

- Extract *statistically significant* patterns

Graph Representation Learning

- Learn low-dimensional *embedding* of nodes

Graph Generative Models



Original Graph H

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Definitions

Graph Models

Overview

String Grammars

Graph Grammars

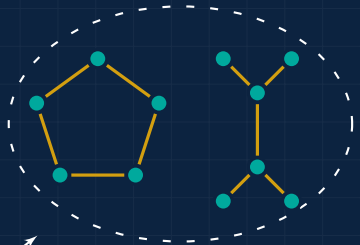
Grammar Extraction

Graph Generation

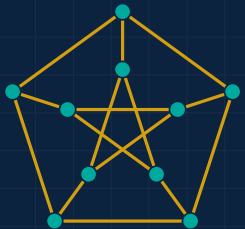
Infinity Mirror Test

Wrapping Up

Graph Generative Models



Extracted Features Θ



Original Graph H

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Definitions

Graph Models

Overview

String Grammars

Graph Grammars

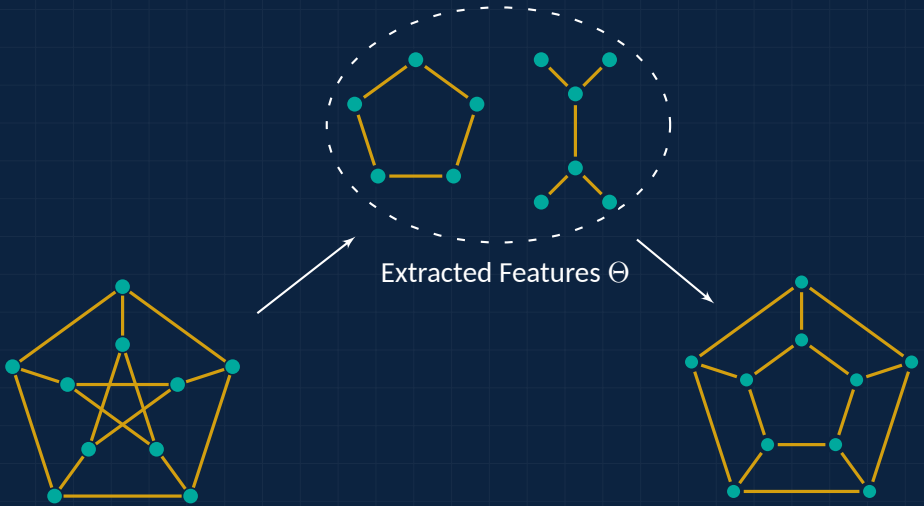
Grammar Extraction

Graph Generation

Infinity Mirror Test

Wrapping Up

Graph Generative Models

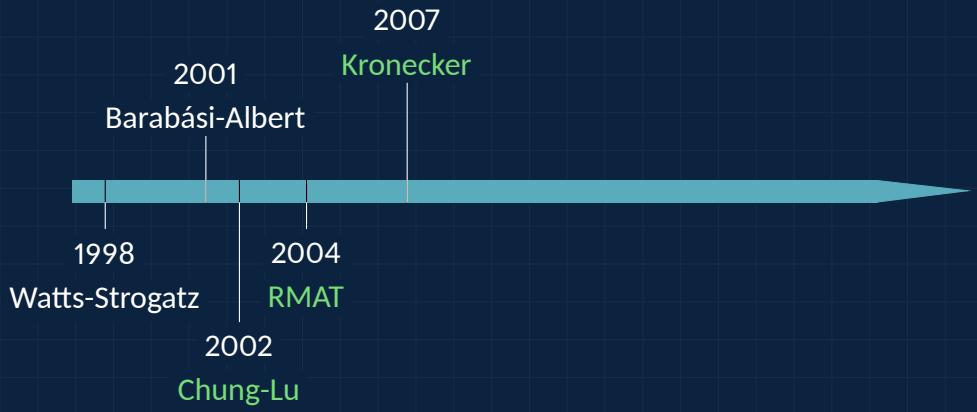


Introduction
Definitions
Graph Models
Overview
String Grammars
Graph Grammars
Grammar Extraction
Graph Generation
Infinity Mirror Test
Wrapping Up

A Brief History of Recent Graph Models

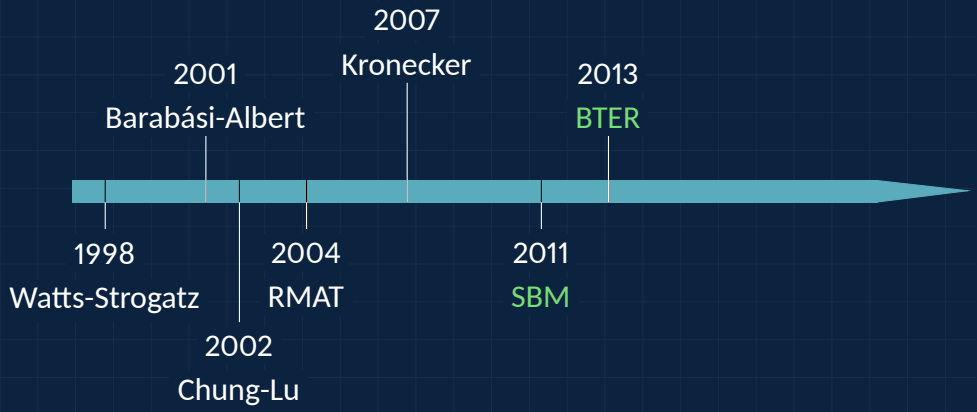
Introduction
Definitions
Graph Models
Overview
String Grammars
Graph Grammars
Grammar Extraction
Graph Generation
Infinity Mirror Test
Wrapping Up

A Brief History of Recent Graph Models

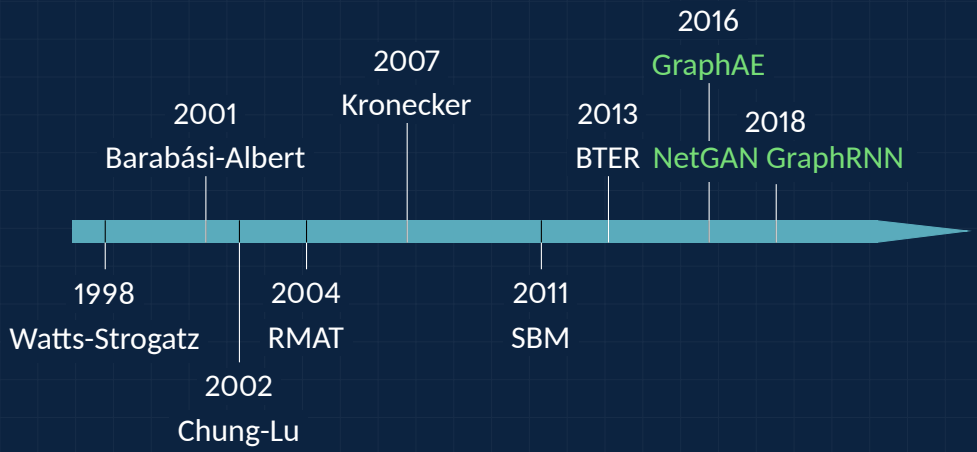


Introduction
Definitions
Graph Models
Overview
String Grammars
Graph Grammars
Grammar Extraction
Graph Generation
Infinity Mirror Test
Wrapping Up

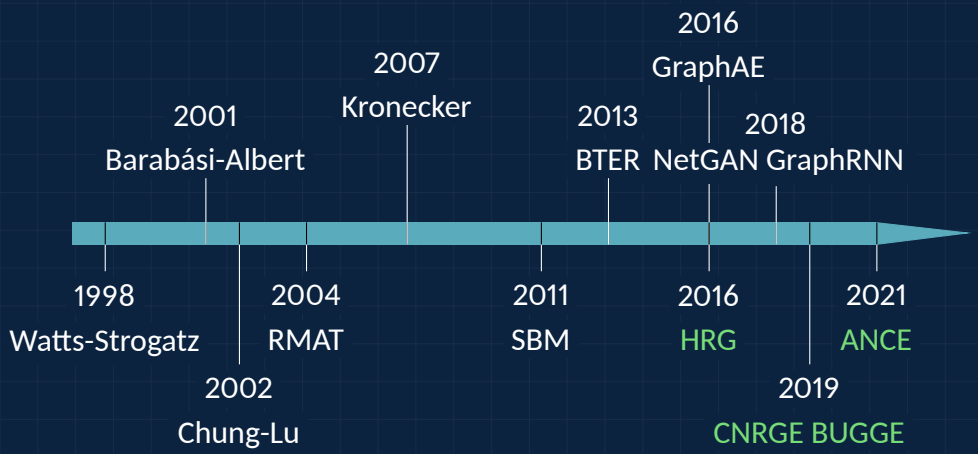
A Brief History of Recent Graph Models



A Brief History of Recent Graph Models



A Brief History of Recent Graph Models



A Fork in the Road

Thesis Statement

Scalably and methodically extracting graph grammar rules provide a unique insight into understanding the inner workings of real-world graphs. Furthermore, these rules can generate accurate copies of the input, allowing for highly interpretable models.

Analyzing the performance of graph models undergoing a stress test reveals and amplifies the latent biases and sheds new light on their inner workings.

A Fork in the Road

List of Publications

ICGT 18	*Synchronous Hyperedge Replacement Graph Grammars*
ICDM 19	*Modeling Graphs with Vertex Replacement Grammars*
BigData 19	*Towards Interpretable Graph Modeling with Vertex Replacement Grammars*
WSDM 21	*Joint Subgraph-to-Subgraph Transitions: Generalizing Triadic Closure for Powerful and Interpretable Graph Modeling*
TKDE 21	*The Infinity Mirror Test for Graph Models*
WSDM 22+	*Attributed Vertex Replacement Graph Grammars*

Introduction
Definitions
Graph Models
Overview
String Grammars
Graph Grammars
Grammar Extraction
Graph Generation
Infinity Mirror Test
Wrapping Up

A Fork in the Road

List of Publications

ICCT 18 *Synchronous Hyperedge Replacement Graph Grammars*

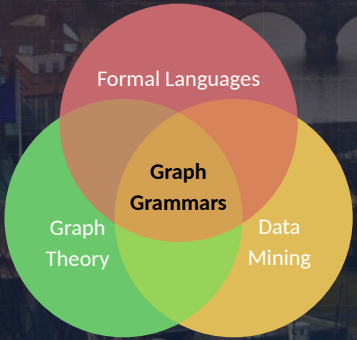
ICDM 19 *Modeling Graphs with Vertex Replacement Grammars*

BigData 19 *Towards Interpretable Graph Modeling with Vertex Replacement Grammars*

WSDM 21 *Joint Subgraph-to-Subgraph Transitions: Generalizing Triadic Closure for Powerful and Interpretable Graph Modeling*

TKDE 21 *The Infinity Mirror Test for Graph Models*

WSDM 22∗ *Attributed Vertex Replacement Graph Grammars*



ICDM 19 *Modeling Graphs with Vertex Replacement Grammars*

WSDM 22∗ *Attributed Vertex Replacement Graph Grammars*

The Big Picture

Studying Complex Systems

- Discovering the *building blocks*
- Figuring how the pieces *fit* together

Formalism

- String grammars for natural *language* processing
- Graph grammars for *graph* mining

String Grammars

CFG Production Rules

- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats

String Grammars

CFG Production Rules

- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats

S

S

String Grammars

CFG Production Rules

- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats

S

S

String Grammars

CFG Production Rules

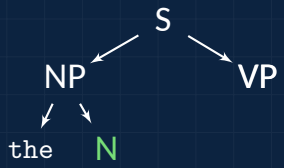
- S $\rightarrow$ NP VP
- NP $\rightarrow$ the N
- VP $\rightarrow$ V NP
- N $\rightarrow$ cat | song
- V $\rightarrow$ sings | eats

$S \Rightarrow NP\ VP$

String Grammars

CFG Production Rules

- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats

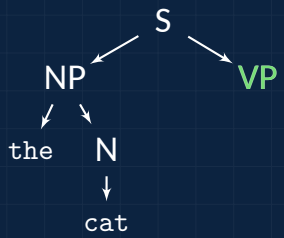


S ⇒ NP VP ⇒ the N VP

String Grammars

CFG Production Rules

- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats

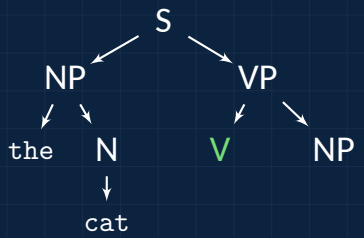


S ⇒ NP VP ⇒ the N VP ⇒ the cat VP

String Grammars

CFG Production Rules

- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats

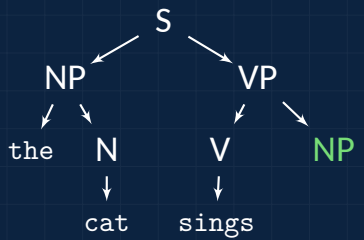


S ⇒ NP VP ⇒ the N VP ⇒ the cat VP ⇒ the cat V NP

String Grammars

CFG Production Rules

- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats

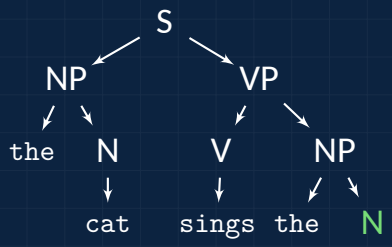


S ⇒ NP VP ⇒ the N VP ⇒ the cat VP ⇒ the cat V NP
⇒ the cat sings NP

String Grammars

CFG Production Rules

- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats

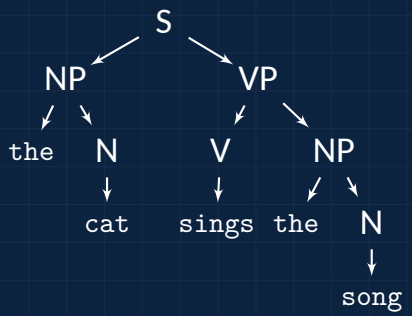


S ⇒ NP VP ⇒ the N VP ⇒ the cat VP ⇒ the cat V NP
⇒ the cat sings NP ⇒ the cat sings the N

String Grammars

CFG Production Rules

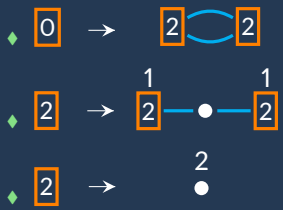
- S → NP VP
- NP → the N
- VP → V NP
- N → cat | song
- V → sings | eats



$$S \Rightarrow NP\ VP \Rightarrow \text{the}\ N\ VP \Rightarrow \text{the cat}\ VP \Rightarrow \text{the cat}\ V\ NP$$
$$\Rightarrow \text{the cat sings}\ NP \Rightarrow \text{the cat sings the}\ N$$
$$\Rightarrow \text{the cat sings the song}$$

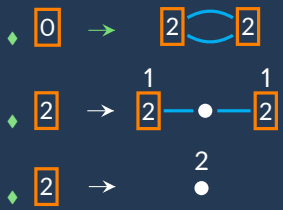
Context Free Graph Grammars

NCE Production Rules



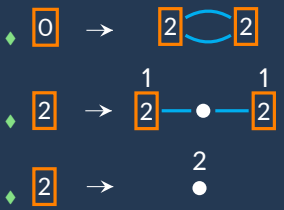
Context Free Graph Grammars

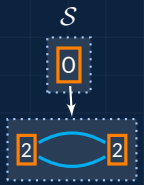
NCE Production Rules

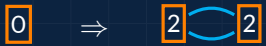


Context Free Graph Grammars

NCE Production Rules

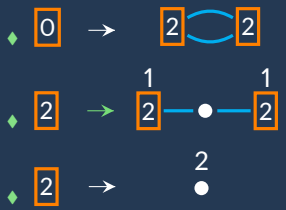


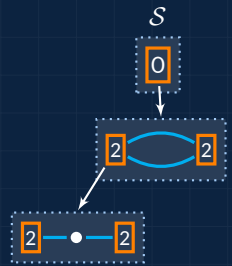


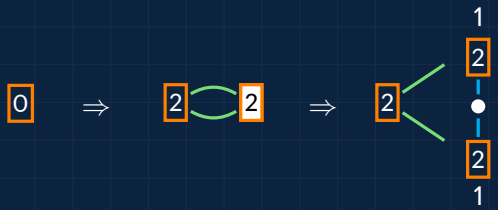


Context Free Graph Grammars

NCE Production Rules







Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Definitions

Graph Models

Overview

String Grammars

Graph Grammars

Grammar Extraction

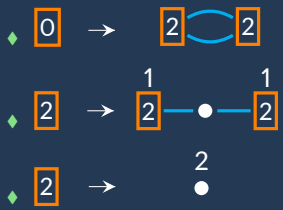
Graph Generation

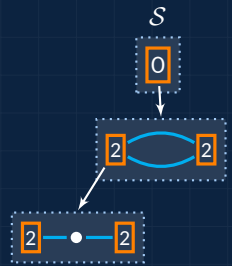
Infinity Mirror Test

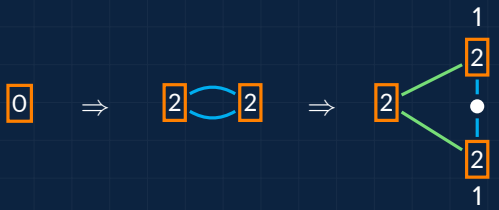
Wrapping Up

Context Free Graph Grammars

NCE Production Rules

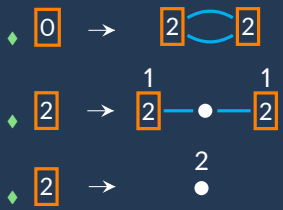


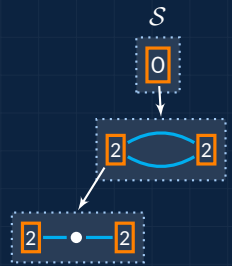


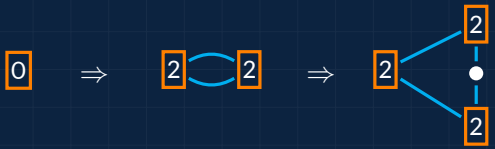


Context Free Graph Grammars

NCE Production Rules

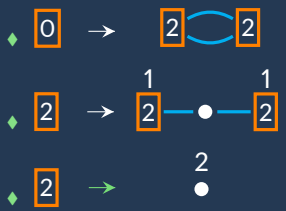


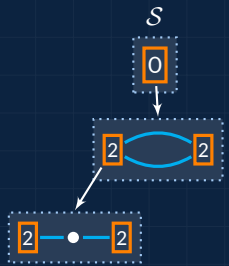




Context Free Graph Grammars

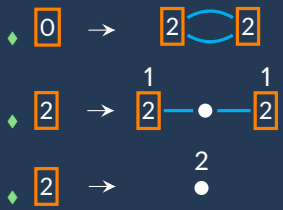
NCE Production Rules

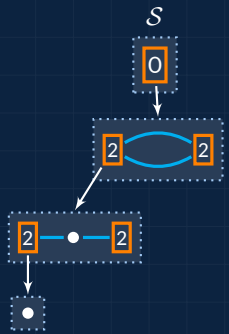




Context Free Graph Grammars

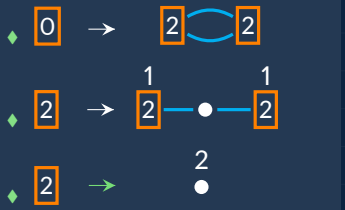
NCE Production Rules

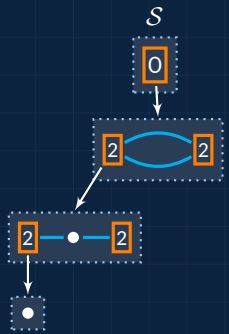




Context Free Graph Grammars

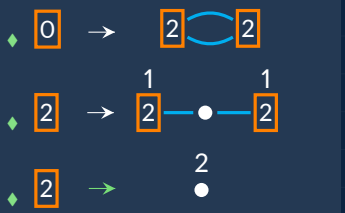
NCE Production Rules

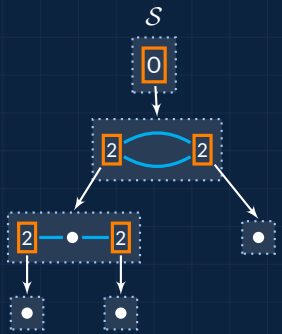


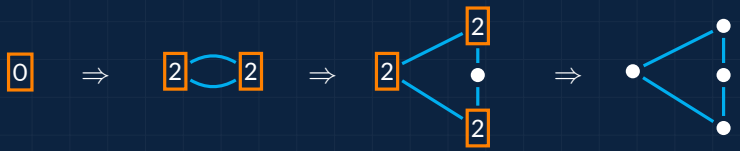


Context Free Graph Grammars

NCE Production Rules







Graph Grammar Pipeline

Input Graph H

Introduction

Definitions

Graph Models

Overview

String Grammars

Graph Grammars

Grammar Extraction

Graph Generation

Infinity Mirror Test

Wrapping Up

Graph Grammar Pipeline

Input Graph H

Grammar Extractor

Graph Grammar Pipeline

Input Graph H

Bottom-up, Clustering-based

Graph Grammar Pipeline

Input Graph *H*

Grammar Extractor

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Definitions

Graph Models

Overview

String Grammars

Graph Grammars

Grammar Extraction

Graph Generation

Infinity Mirror Test

Wrapping Up

Graph Grammar Pipeline

Input Graph *H*

Grammar Extractor

Graph Generator

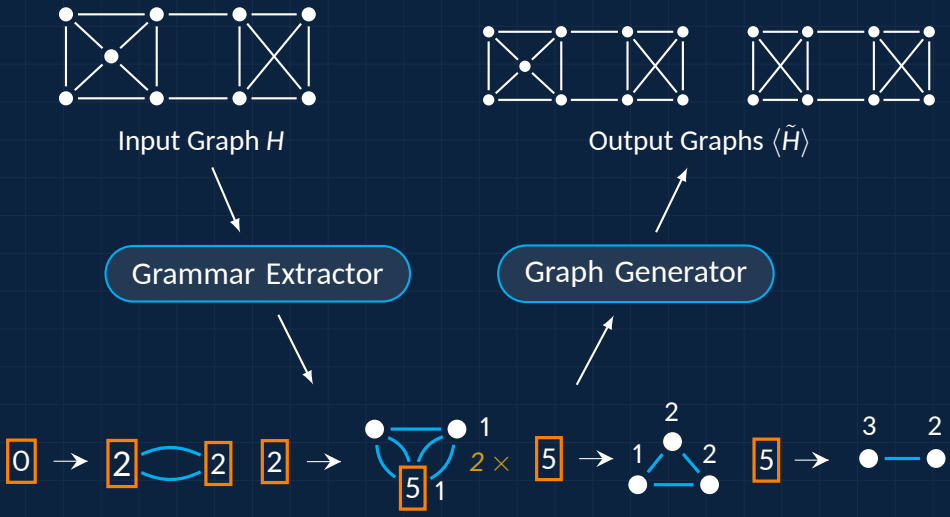
Graph Grammar Pipeline

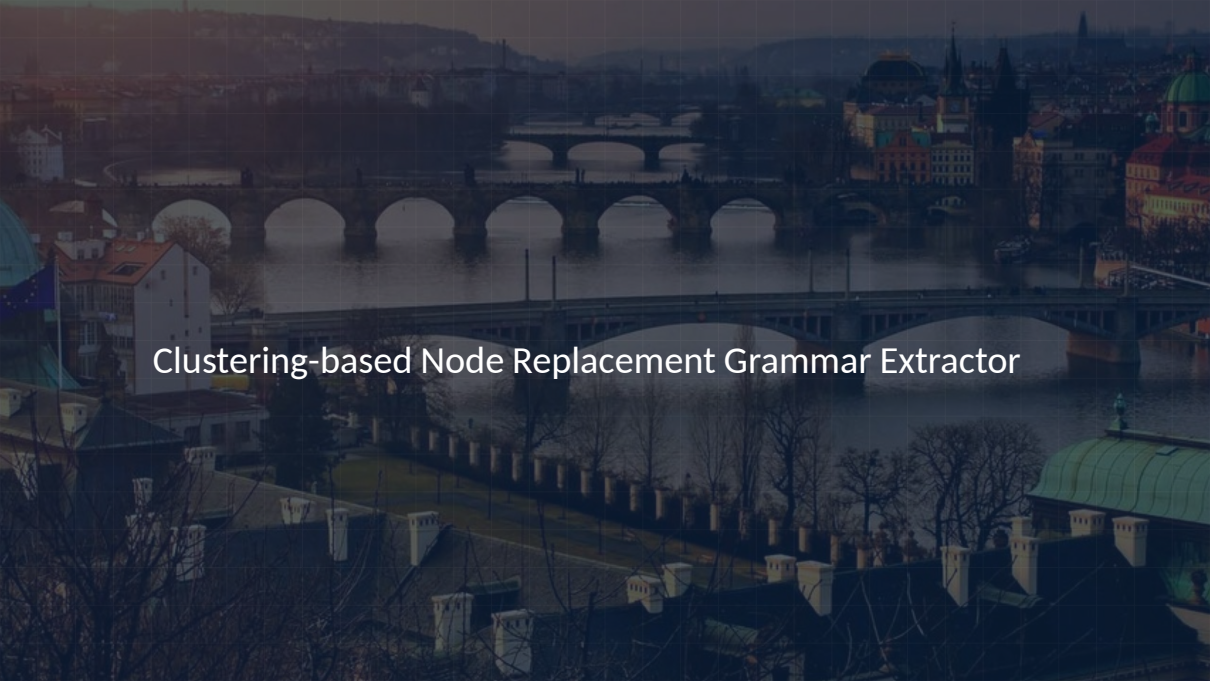
Input Graph *H*

Grammar Extractor

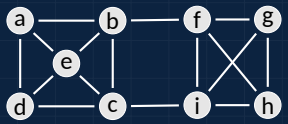
Random, Attribute-aware

Graph Grammar Pipeline




Clustering-based Node Replacement Grammar Extractor

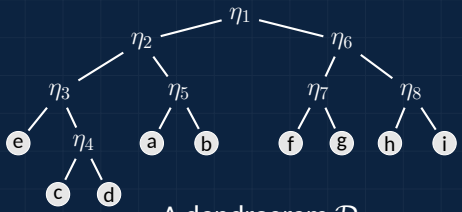
Hierarchical Graph Clustering



Graph H

Leiden
Louvain

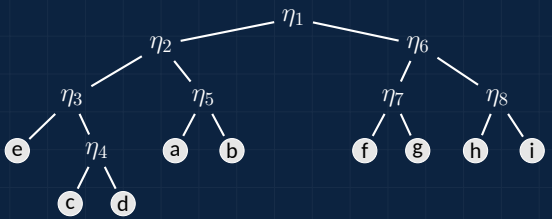
Spectral
MinCut



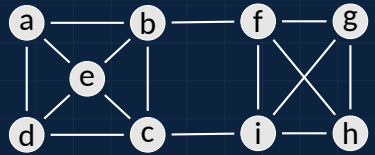
A dendrogram $\mathcal{D}$

Extracting NCE Rules I

($\mu = 3$)



Current Dendrogram $\mathcal{D}$



Current Graph H

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

Attributed Graphs

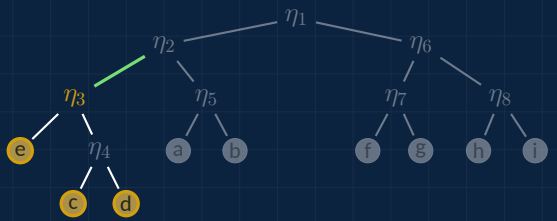
Graph Generation

Infinity Mirror Test

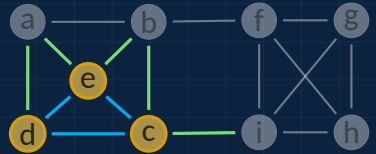
Wrapping Up

Extracting NCE Rules I

($\mu = 3$)



Current Dendrogram $\mathcal{D}$



Current Graph H

Extracted NCE Rule

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

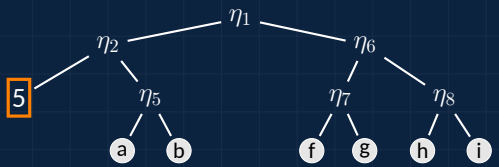
Attributed Graphs

Graph Generation

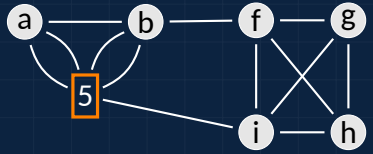
Infinity Mirror Test

Wrapping Up

Extracting NCE Rules I ($\mu = 3$)



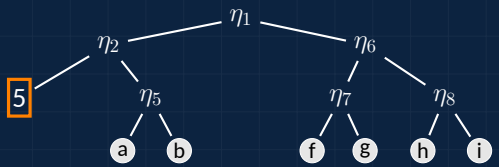
Updated Dendrogram $\mathcal{D}'$



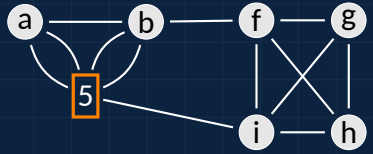
Updated Graph H'

Extracting NCE Rules II

($\mu = 3$)



Current Dendrogram $\mathcal{D}$



Current Graph H

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

Attributed Graphs

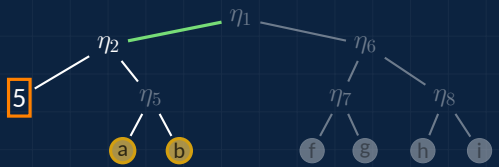
Graph Generation

Infinity Mirror Test

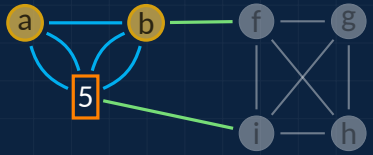
Wrapping Up

Extracting NCE Rules II

($\mu = 3$)



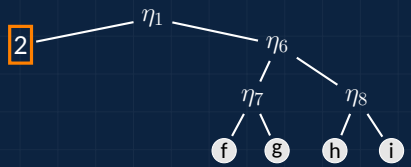
Current Dendrogram $\mathcal{D}$



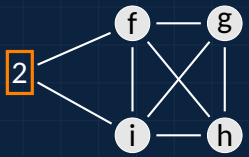
Current Graph H

Extracted NCE Rule

Extracting NCE Rules II ($\mu = 3$)



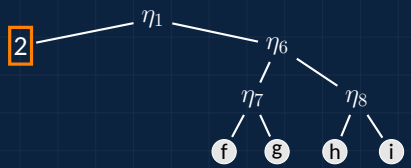
Updated Dendrogram $\mathcal{D}'$



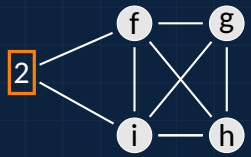
Updated Graph H'

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Motivation
Rule Extraction
Attributed Graphs
Graph Generation
Infinity Mirror Test
Wrapping Up

Extracting NCE Rules III ($\mu = 3$)



Current Dendrogram $\mathcal{D}$



Current Graph H

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

Attributed Graphs

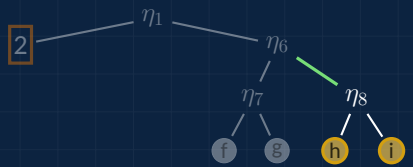
Graph Generation

Infinity Mirror Test

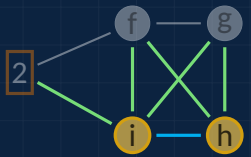
Wrapping Up

Extracting NCE Rules III

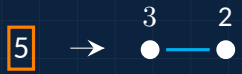
($\mu = 3$)



Current Dendrogram $\mathcal{D}$



Current Graph H



Extracted NCE Rule

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

Attributed Graphs

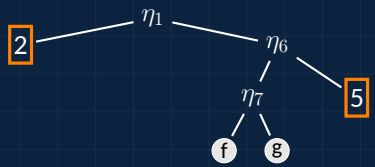
Graph Generation

Infinity Mirror Test

Wrapping Up

Extracting NCE Rules III

($\mu = 3$)

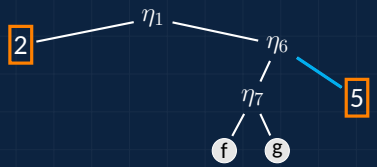


Updated Dendrogram $\mathcal{D}'$

Updated Graph H'

Extracting NCE Rules IV

($\mu = 3$)



Current Dendrogram $\mathcal{D}$

Current Graph *H*

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

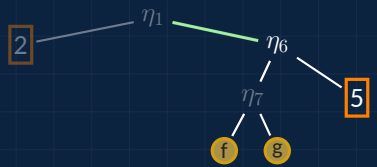
Attributed Graphs

Graph Generation

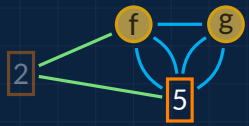
Infinity Mirror Test

Wrapping Up

Extracting NCE Rules IV ($\mu = 3$)



Current Dendrogram $\mathcal{D}$



Current Graph H

Extracted NCE Rule

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

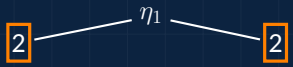
Attributed Graphs

Graph Generation

Infinity Mirror Test

Wrapping Up

Extracting NCE Rules IV ($\mu = 3$)



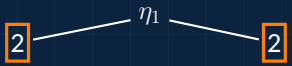
Updated Dendrogram $\mathcal{D}'$

Updated Graph H'

Extracted NCE Rule

Extracting NCE Rules V

($\mu = 3$)

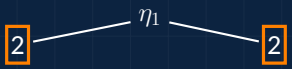


Current Dendrogram $\mathcal{D}$

Current Graph H

Extracting NCE Rules V

($\mu = 3$)



Current Dendrogram $\mathcal{D}$

Current Graph H

Extracted NCE Rule

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

Attributed Graphs

Graph Generation

Infinity Mirror Test

Wrapping Up

Extracting NCE Rules V

($\mu = 3$)

Updated Dendrogram $\mathcal{D}$

Updated Graph H

Extracting NCE Rules V

($\mu = 3$)

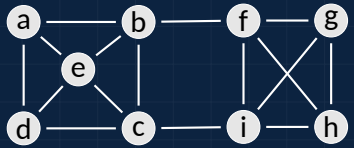
0

0

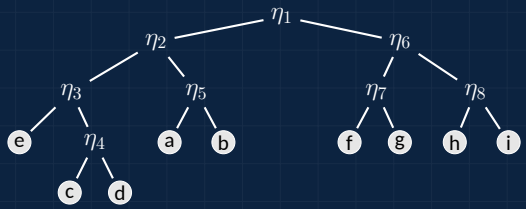
Updated Dendrogram $\mathcal{D}$

Updated Graph H

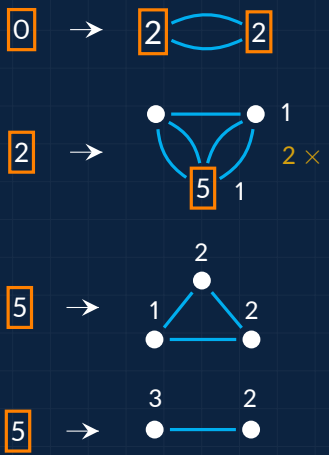
Extracted NCE Grammar



Graph H



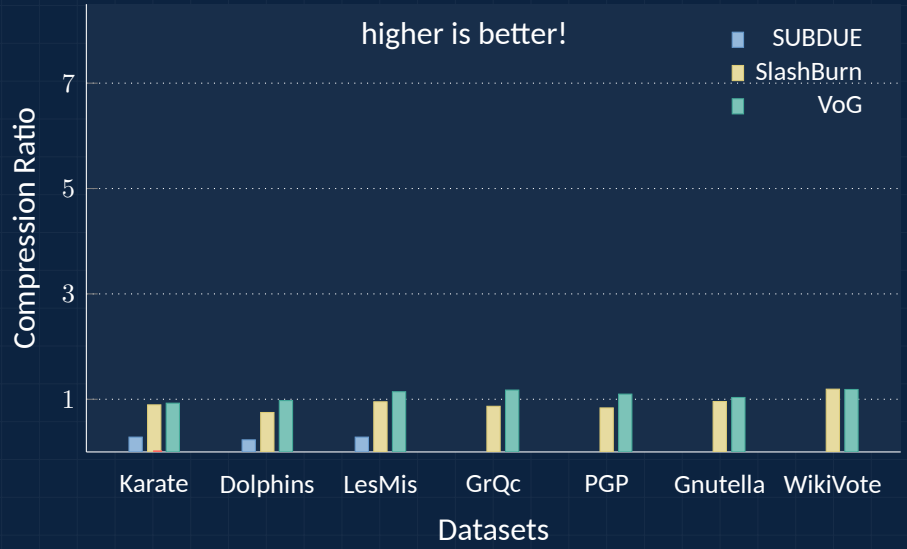
Dendrogram $\mathcal{D}$



Extracted NCE rules

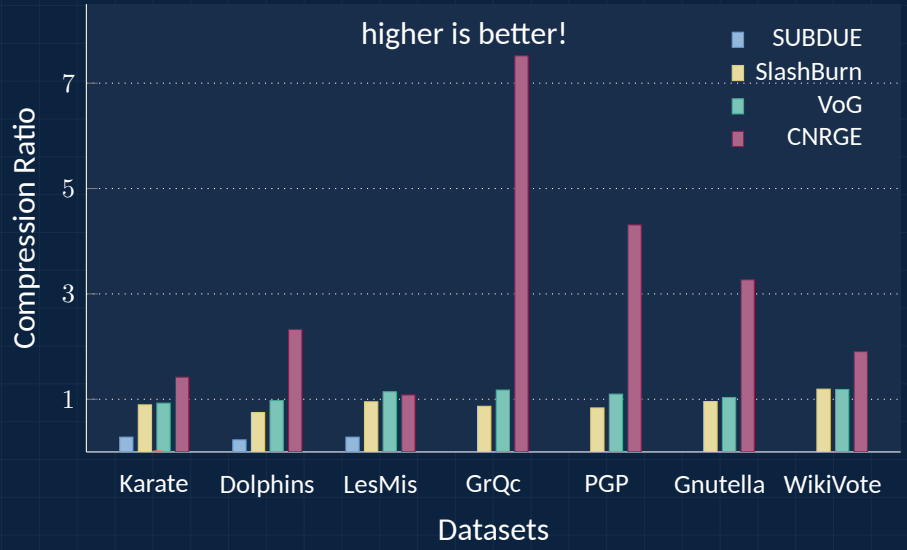
Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Motivation
Rule Extraction
Attributed Graphs
Graph Generation
Infinity Mirror Test
Wrapping Up

Model Size Comparison

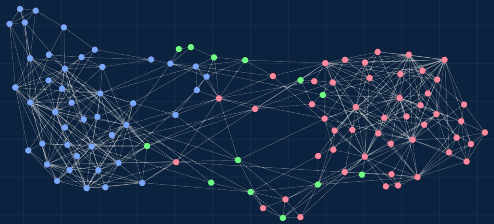


Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Motivation
Rule Extraction
Attributed Graphs
Graph Generation
Infinity Mirror Test
Wrapping Up

Model Size Comparison



Attributed Graphs



Homophily in PolBooks graph

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

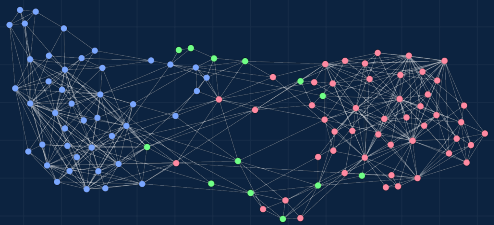
Attributed Graphs

Graph Generation

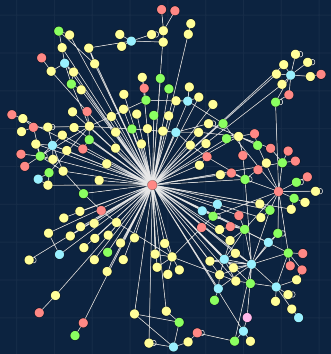
Infinity Mirror Test

Wrapping Up

Attributed Graphs



Homophily in PolBooks graph



Heterophily in Texas CS website

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Motivation

Rule Extraction

Attributed Graphs

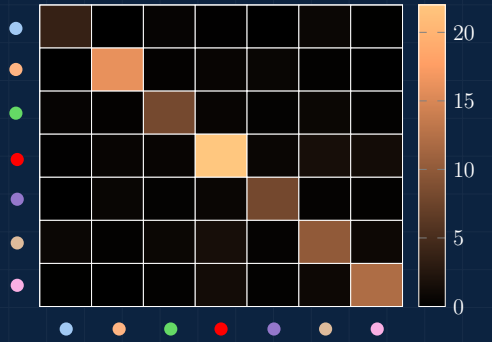
Graph Generation

Infinity Mirror Test

Wrapping Up

Real-world Grammars: Cora

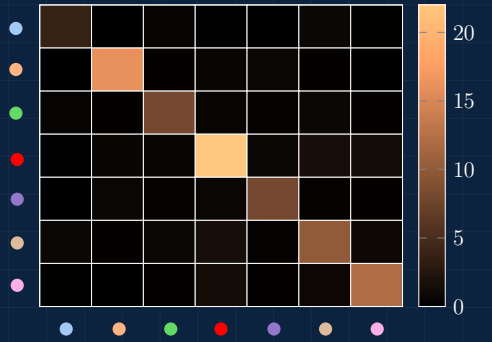
Node Distribution



Edge Mixing Matrix

Real-world Grammars: Cora

Node Distribution



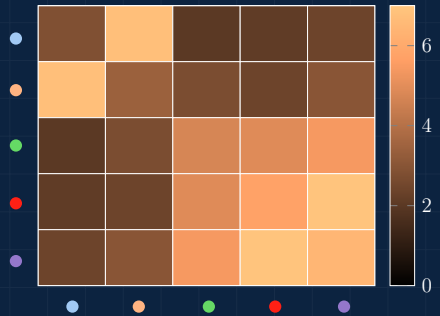
Edge Mixing Matrix

6 Most-frequent Rule RHSs

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Motivation
Rule Extraction
Attributed Graphs
Graph Generation
Infinity Mirror Test
Wrapping Up

Real-world Grammars: Chameleon

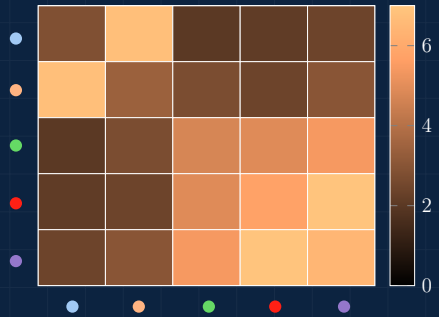
Node Distribution



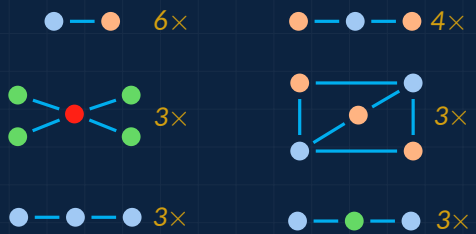
Edge Mixing Matrix

Real-world Grammars: Chameleon

Node Distribution

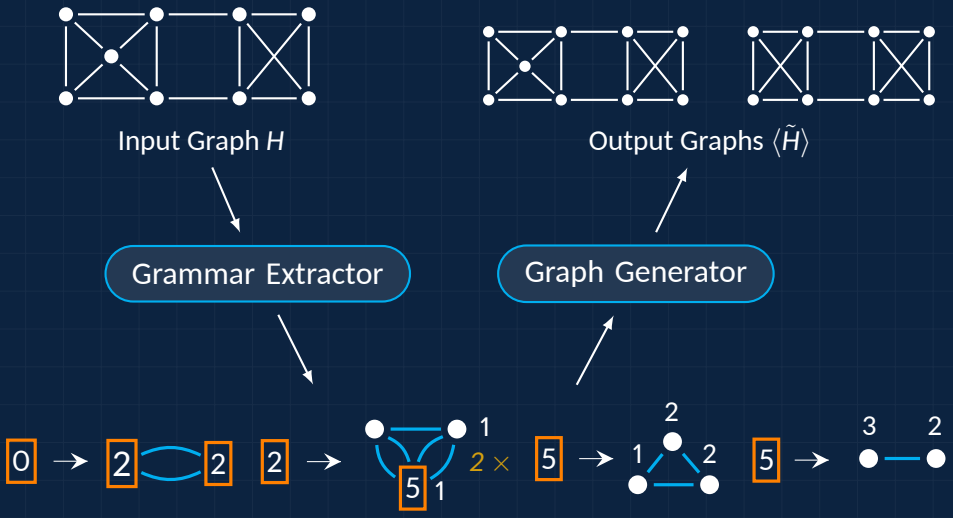


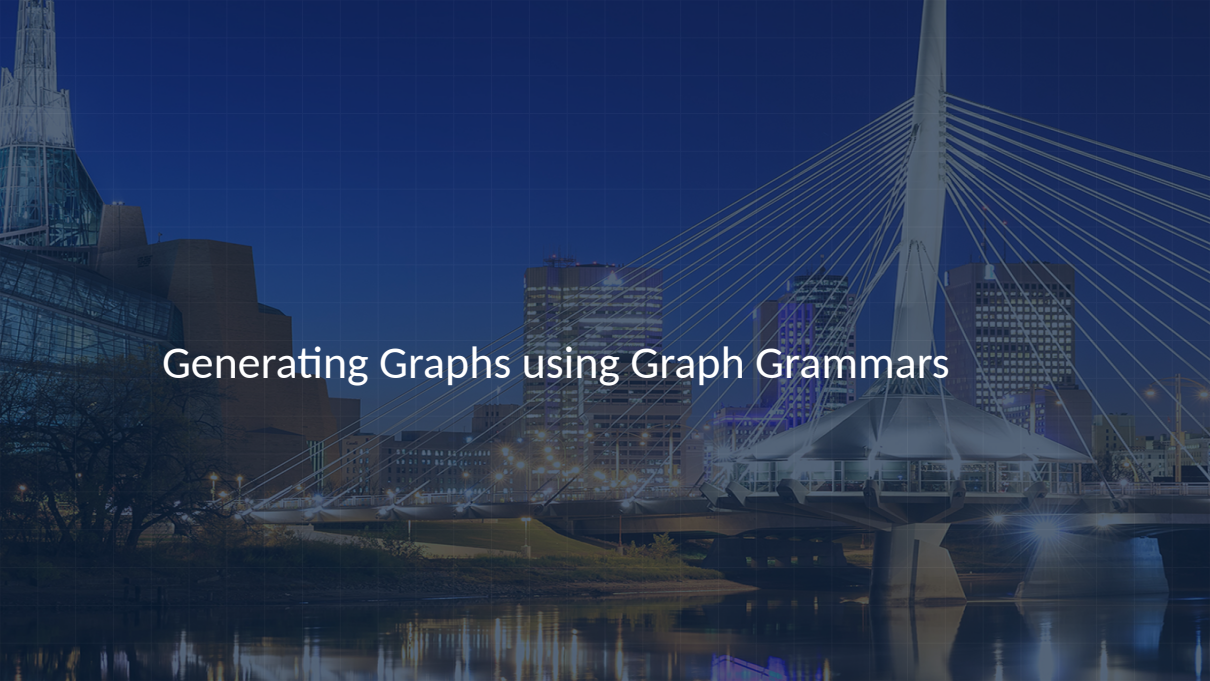
Edge Mixing Matrix



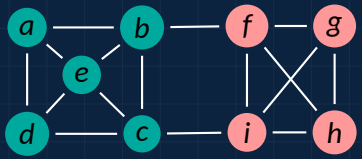
6 Most-frequent Rule RHSs

Graph Grammar Pipeline Revisited

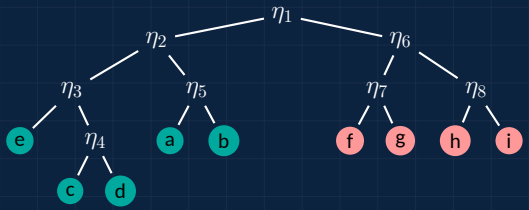



Generating Graphs using Graph Grammars

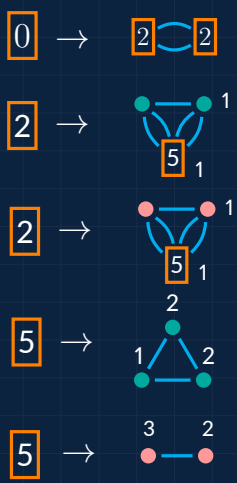
Attributed NCE Grammars



Graph H



Dendrogram $\mathcal{D}$



Extracted ANCE rules

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Methodology

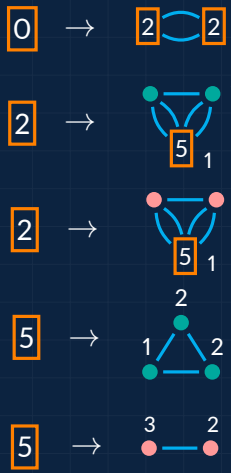
Experimental Results

Conclusions

Infinity Mirror Test

Wrapping Up

Generating Graphs from an ANCE I



Extracted ANCE Rules

$\mathcal{S}$

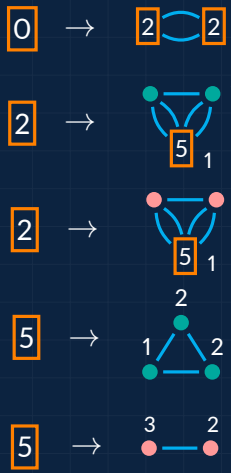
0

Current Graph H'

Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE I



Extracted ANCE Rules

$\mathcal{S}$

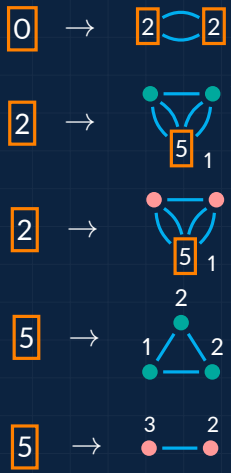
0

Current Graph H'

Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE I



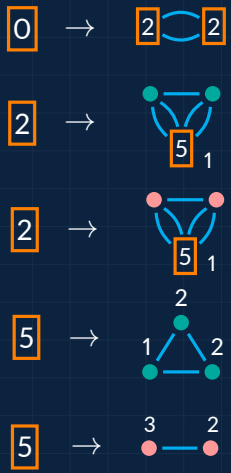
Extracted ANCE Rules

Current Graph H'

Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE II



Extracted ANCE Rules

Current Graph H'

Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Methodology

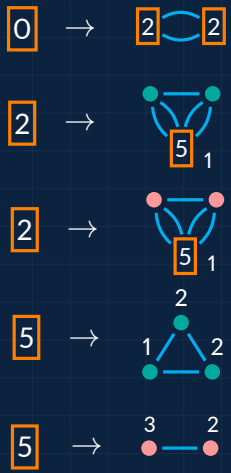
Experimental Results

Conclusions

Infinity Mirror Test

Wrapping Up

Generating Graphs from an ANCE II



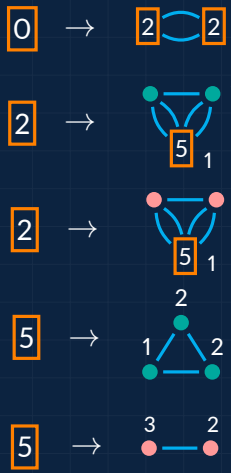
Extracted ANCE Rules

Current Graph H'

Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE II



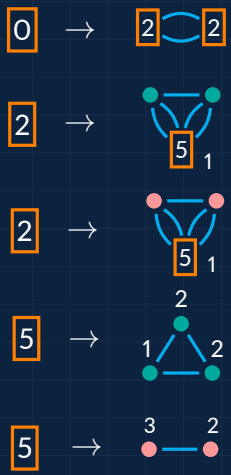
Extracted ANCE Rules

Current Graph H'

Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE III



Extracted ANCE Rules

Current Graph H'

Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Methodology

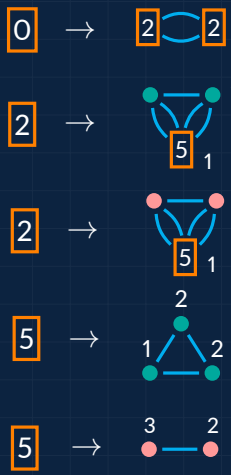
Experimental Results

Conclusions

Infinity Mirror Test

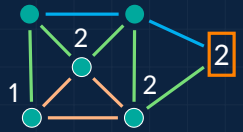
Wrapping Up

Generating Graphs from an ANCE III



Extracted ANCE Rules

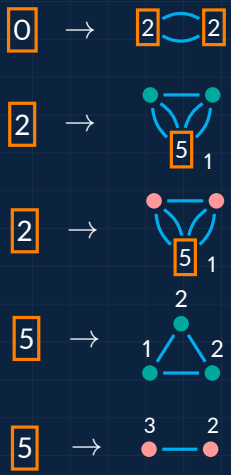
Current Graph H'



Next Graph $\hat{H}$

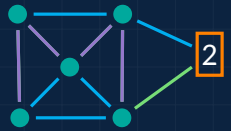
Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE III



Extracted ANCE Rules

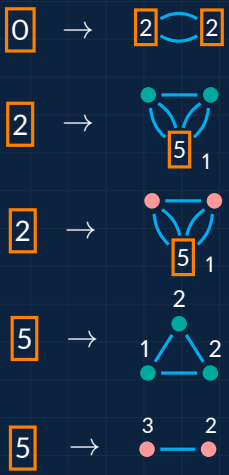
Current Graph H'



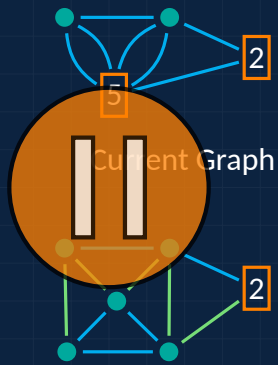
Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE III



Extracted ANCE Rules



Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Methodology

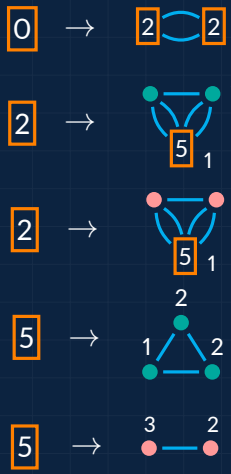
Experimental Results

Conclusions

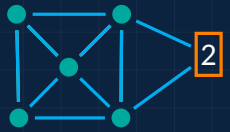
Infinity Mirror Test

Wrapping Up

Generating Graphs from an ANCE IV



Extracted ANCE Rules

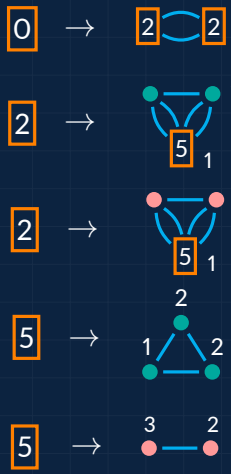


Current Graph H'

Next Graph $\hat{H}$

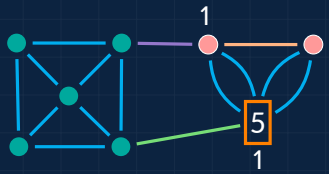
Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE IV



Extracted ANCE Rules

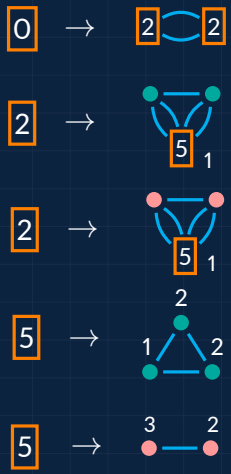
Current Graph H'



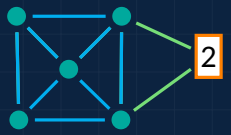
Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE IV



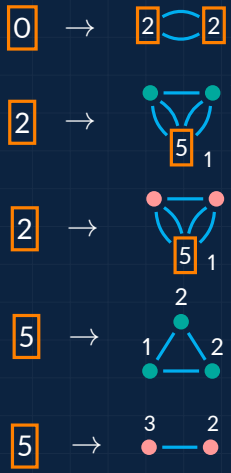
Extracted ANCE Rules



Current Graph H'

Next Graph $\hat{H}$

Generating Graphs from an ANCE V



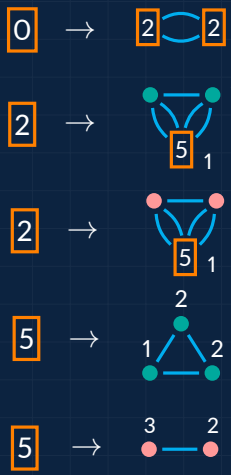
Extracted ANCE Rules

Current Graph H'

Next Graph $\hat{H}$

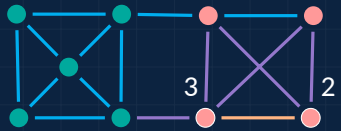
Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE V



Extracted ANCE Rules

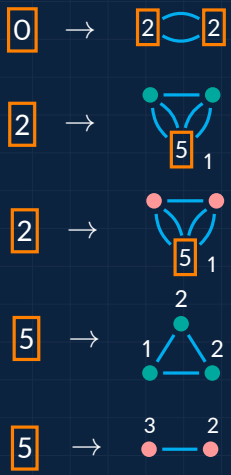
Current Graph H'



Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE V



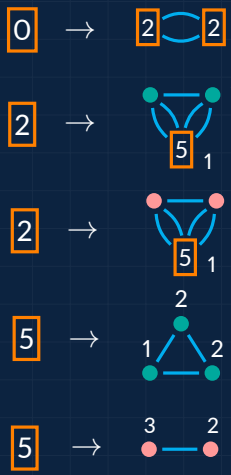
Extracted ANCE Rules

Current Graph H'

Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE V



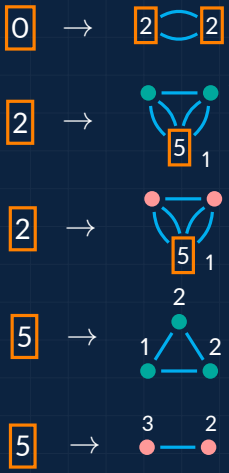
Extracted ANCE Rules

Current Graph H'

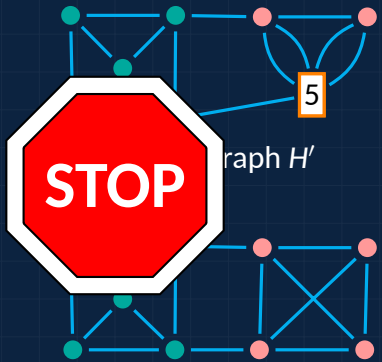
Next Graph $\hat{H}$

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

Generating Graphs from an ANCE V



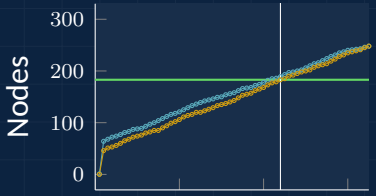
Extracted ANCE Rules

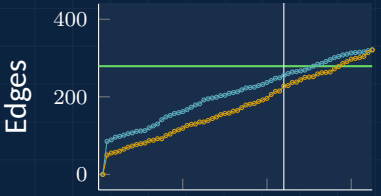


Next Graph $\hat{H}$

Language of Graphs Described by a Grammar

Examining the Graph Growth Process

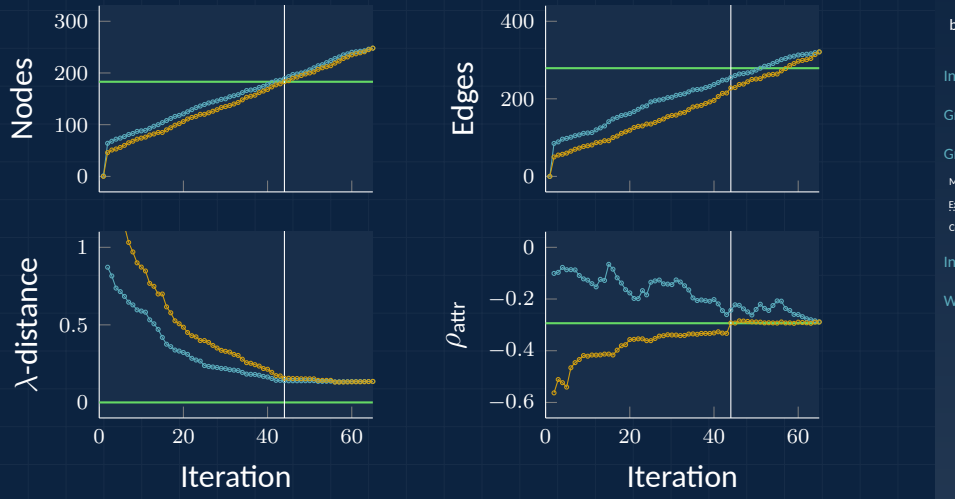




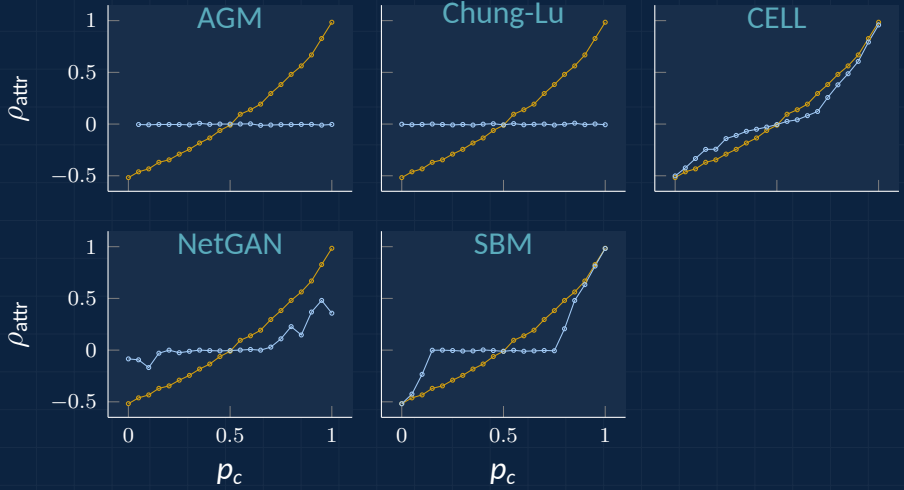
Whole Graph Terminal Graph Input Graph Metric Input Graph Size

Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Methodology
Experimental Results
Conclusions
Infinity Mirror Test
Wrapping Up

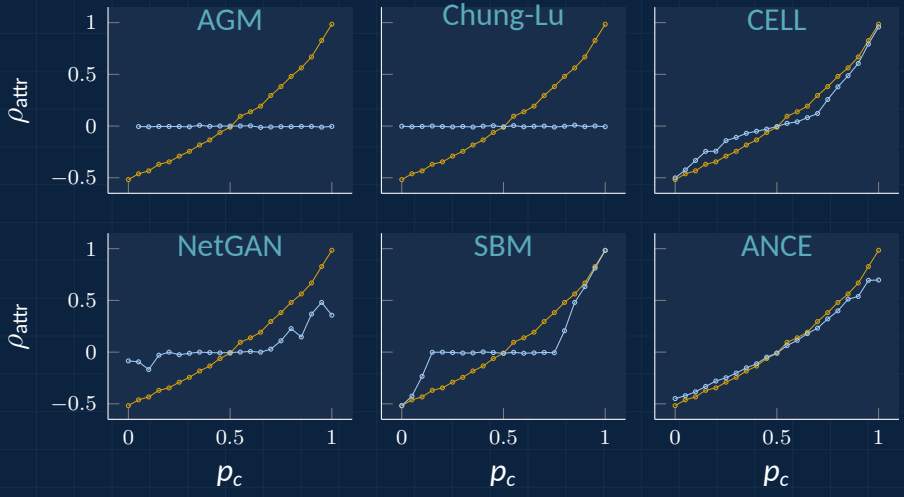
Examining the Graph Growth Process



CABAM Graphs



CABAM Graphs



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Methodology

Experimental Results

Conclusions

Infinity Mirror Test

Wrapping Up

Takeaways

Main Takeaways

- *Simple* and *interpretable* formalism from formal languages

Takeaways

Main Takeaways

- *Simple* and *interpretable* formalism from formal languages
- *Scalable* to medium-large graphs

Takeaways

Main Takeaways

- *Simple* and *interpretable* formalism from formal languages
- *Scalable* to medium-large graphs
- *Faithful* graph generation: topology and attribute

Takeaways

Main Takeaways

- *Simple* and *interpretable* formalism from formal languages
- *Scalable* to medium-large graphs
- *Faithful* graph generation: topology and attribute

Shortcomings

- Dependence on the *dendrogram*

Takeaways

Main Takeaways

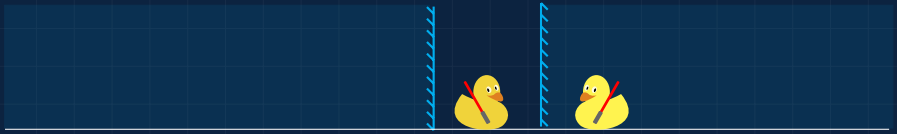
- *Simple* and *interpretable* formalism from formal languages
- *Scalable* to medium-large graphs
- *Faithful* graph generation: topology and attribute

Shortcomings

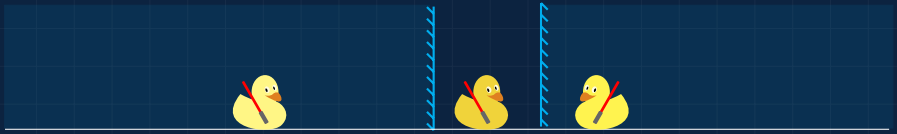
- Dependence on the *dendrogram*
- Rule extraction process is *non-deterministic*

The Infinity Mirror Test for Graph Models

Infinity Mirror Test for Duck Vader



Infinity Mirror Test for Duck Vader



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

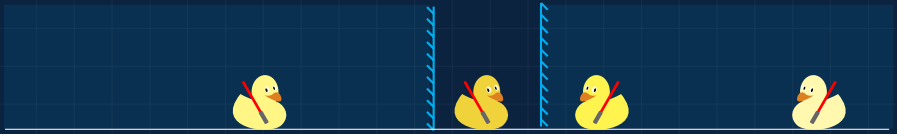
Methodology

Results

Conclusion

Wrapping Up

Infinity Mirror Test for Duck Vader



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

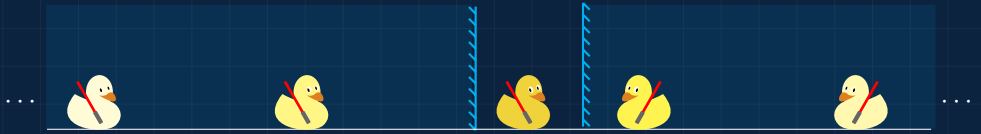
Methodology

Results

Conclusion

Wrapping Up

Infinity Mirror Test for Duck Vader



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

Results

Conclusion

Wrapping Up

Infinity Mirror Test for a Graph

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

Results

Conclusion

Wrapping Up

Infinity Mirror Test for a Graph

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

Results

Conclusion

Wrapping Up

Infinity Mirror Test for a Graph

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

Results

Conclusion

Wrapping Up

Infinity Mirror Test for a Graph

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

Results

Conclusion

Wrapping Up

Infinity Mirror Test for a Graph

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

Results

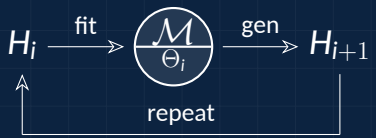
Conclusion

Wrapping Up

Infinity Mirror Test for Graph Models

Key Idea

Forcing a model to fit and re-fit the generated graphs amplifies biases



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

Results

Conclusion

Wrapping Up

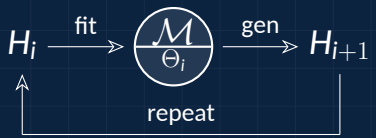
Infinity Mirror Test for Graph Models

Key Idea

Forcing a model to fit and re-fit the generated graphs amplifies biases

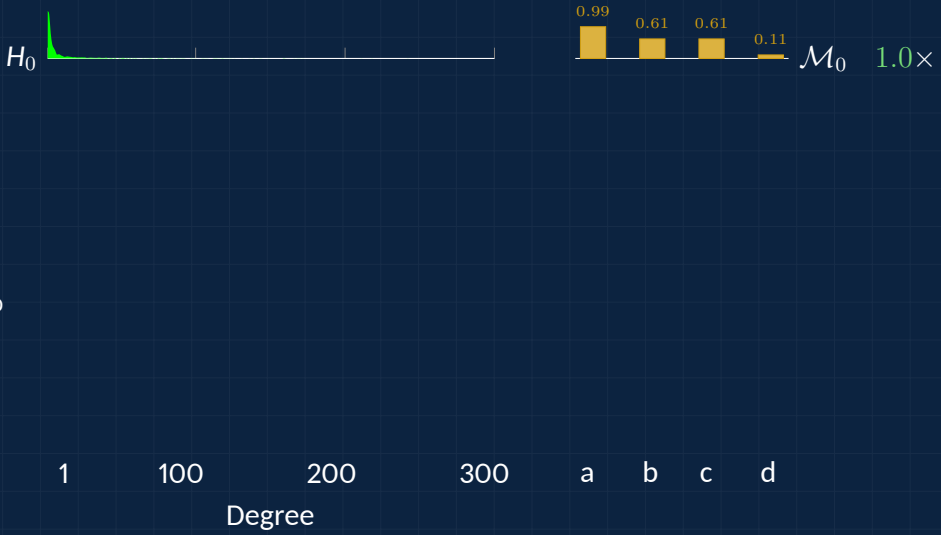
Evaluation Plan

Compare H_i and H_0 to observe *incremental* degradation



Kronecker Degree Distribution Ridge Plot

$\mathcal{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

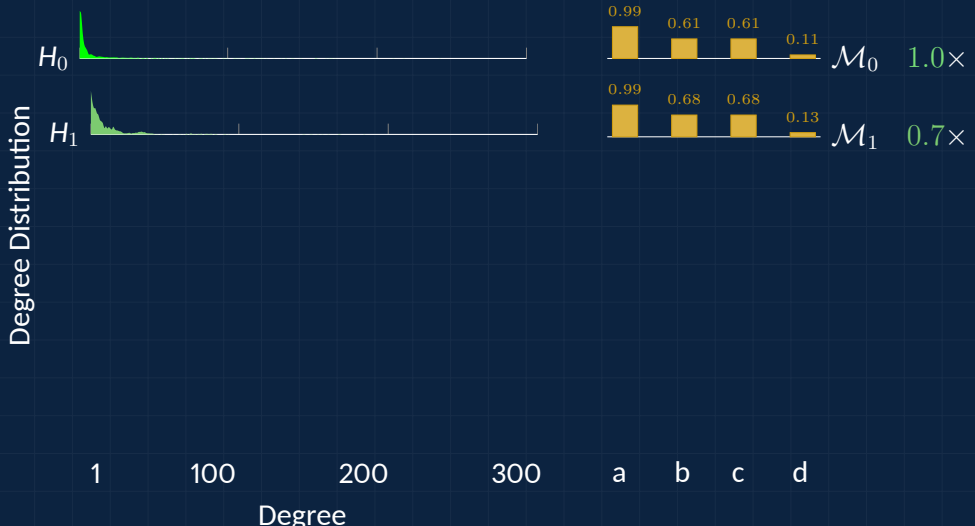
Results

Conclusion

Wrapping Up

Kronecker Degree Distribution Ridge Plot

$\mathcal{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

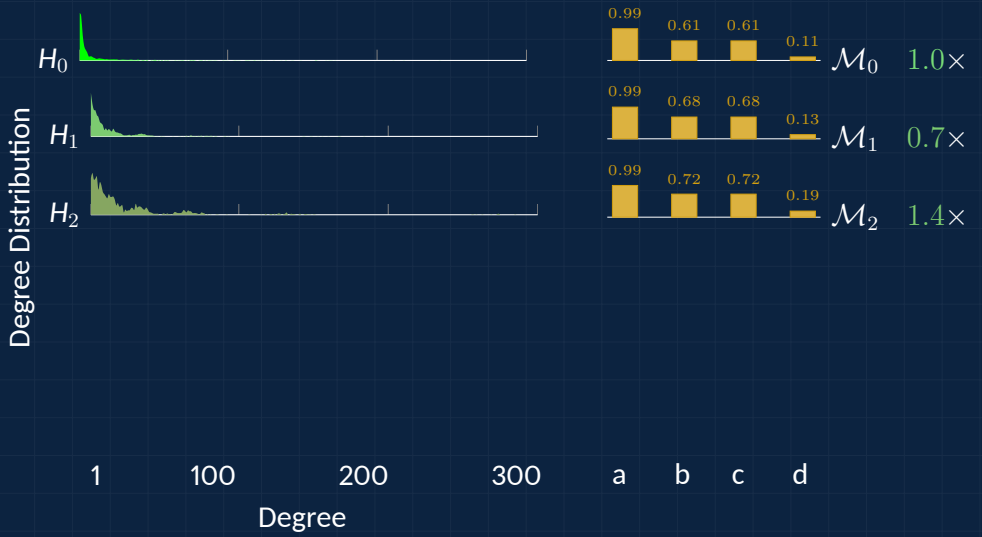
Results

Conclusion

Wrapping Up

Kronecker Degree Distribution Ridge Plot

$\mathcal{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

Methodology

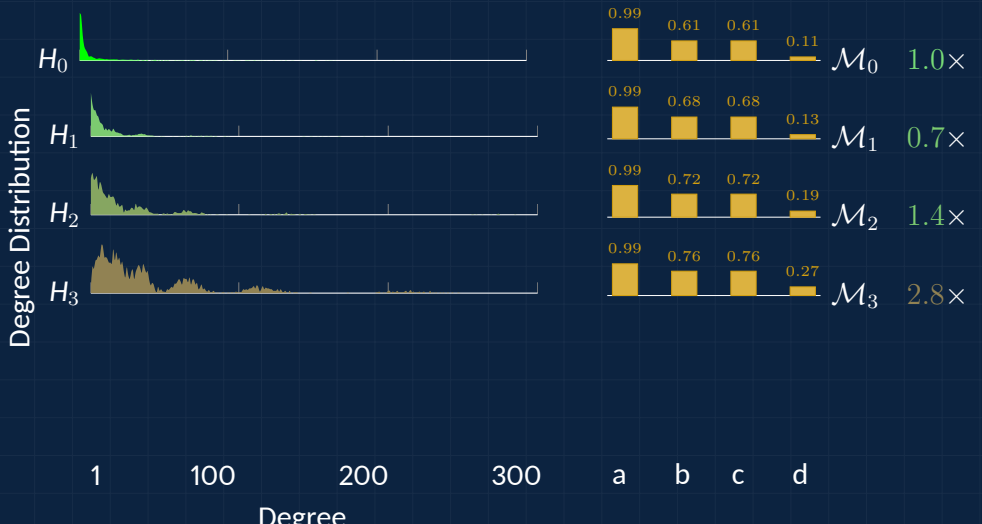
Results

Conclusion

Wrapping Up

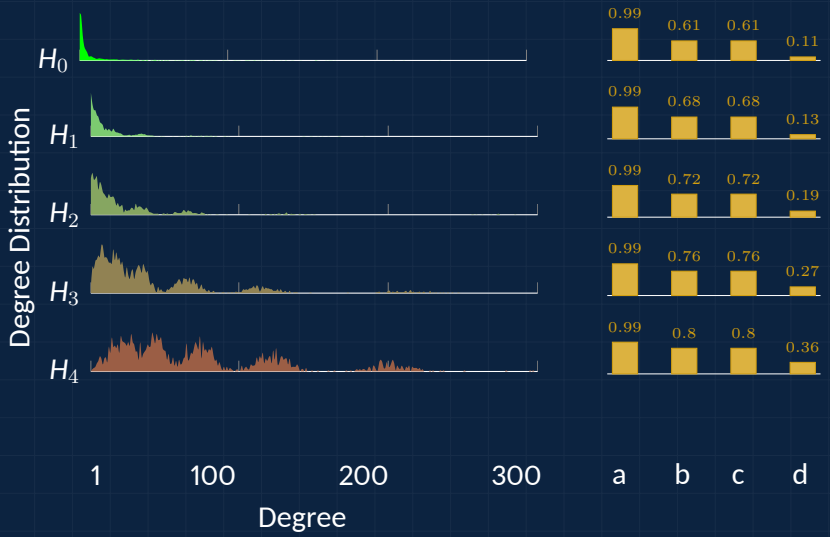
Kronecker Degree Distribution Ridge Plot

$\mathcal{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$



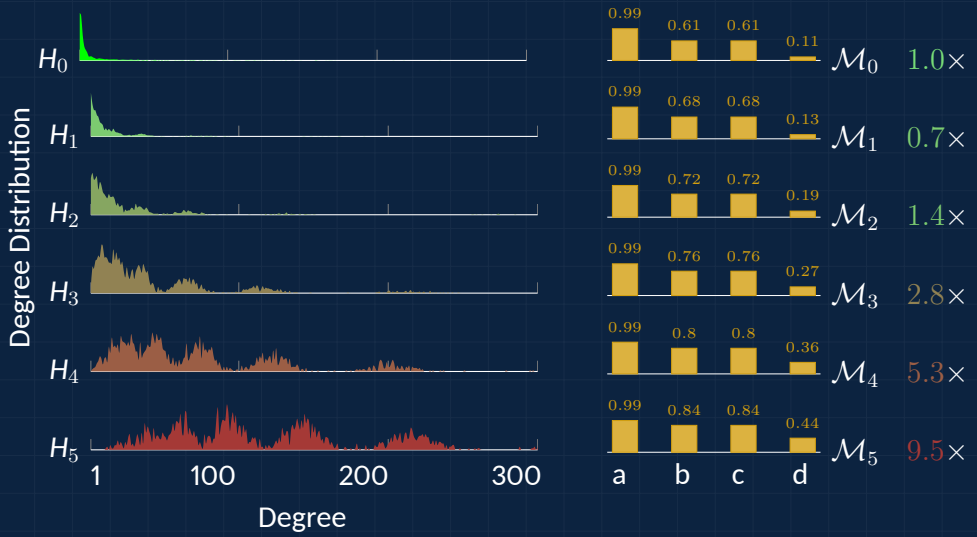
Kronecker Degree Distribution Ridge Plot

$\mathcal{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$



Kronecker Degree Distribution Ridge Plot

$\mathcal{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

Graph Generation

Infinity Mirror Test

Motivation

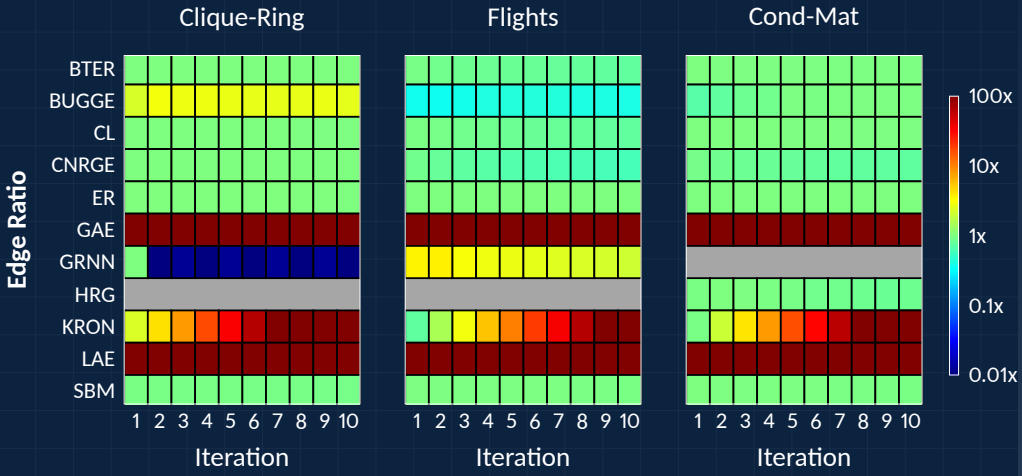
Methodology

Results

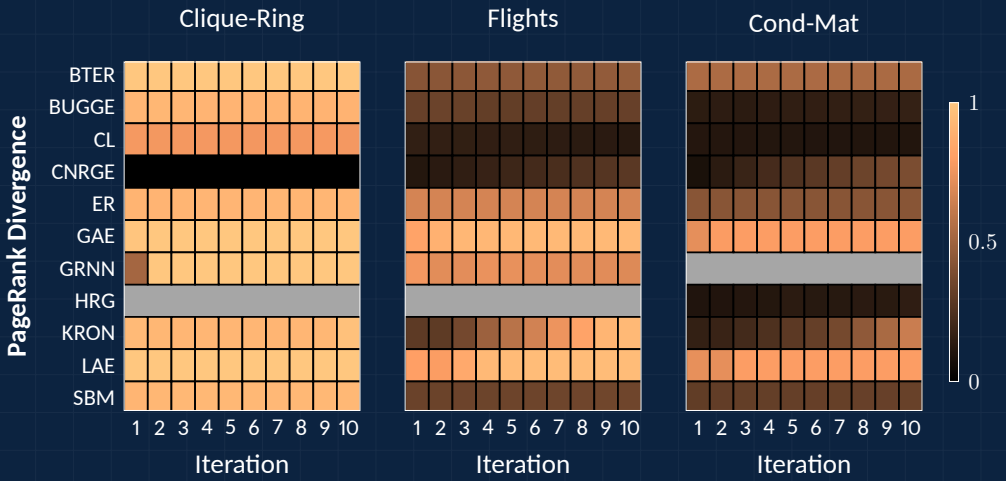
Conclusion

Wrapping Up

First-order Graph Statistics



First-order Graph Statistics



Tracking Evolution: Clustering & Geodesics

Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

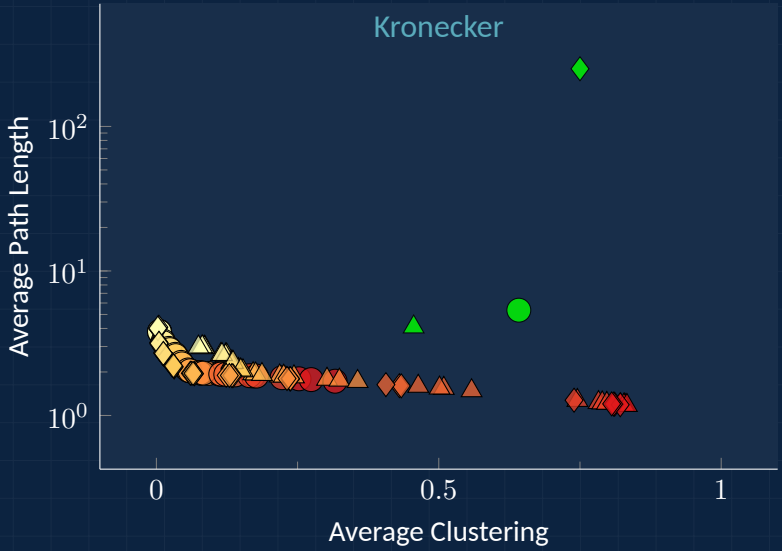
Graph Generation

Infinity Mirror Test

- Motivation
- Methodology
- Results
- Conclusion

Wrapping Up

Tracking Evolution: Clustering & Geodesics



Scalable and Interpretable Graph Modeling with Graph Grammars

bit.ly/satyaki-slides

Introduction

Grammar Extraction

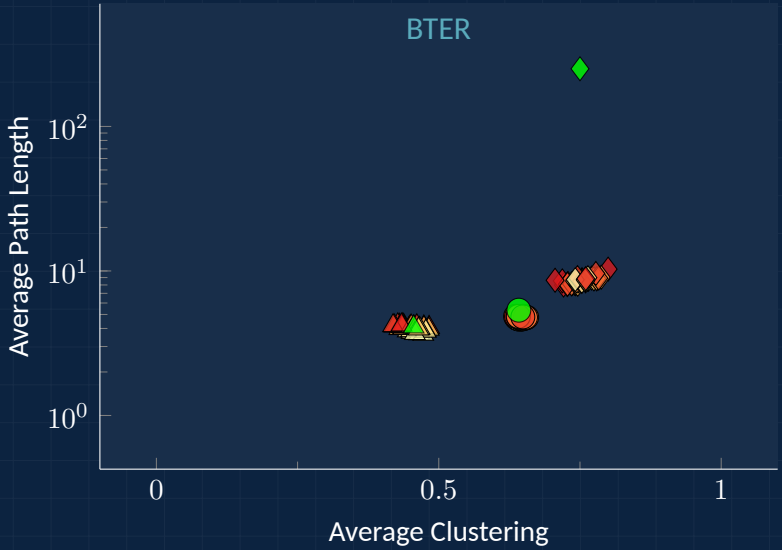
Graph Generation

Infinity Mirror Test

- Motivation
- Methodology
- Results
- Conclusion

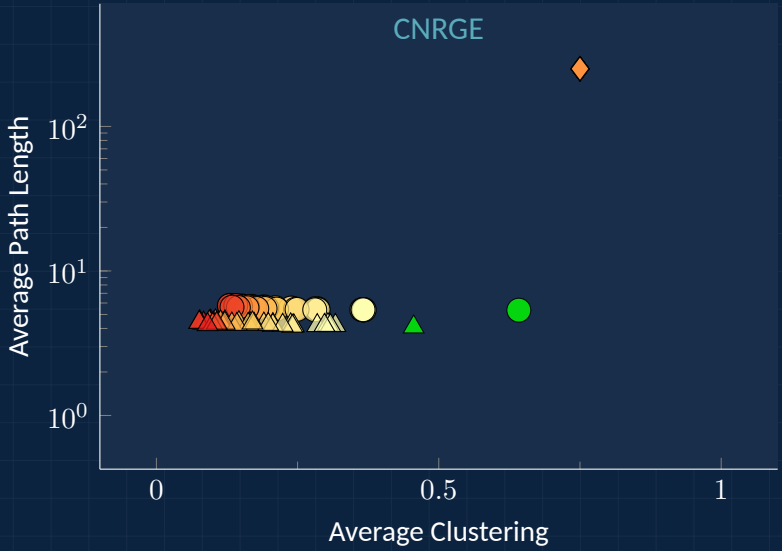
Wrapping Up

Tracking Evolution: Clustering & Geodesics



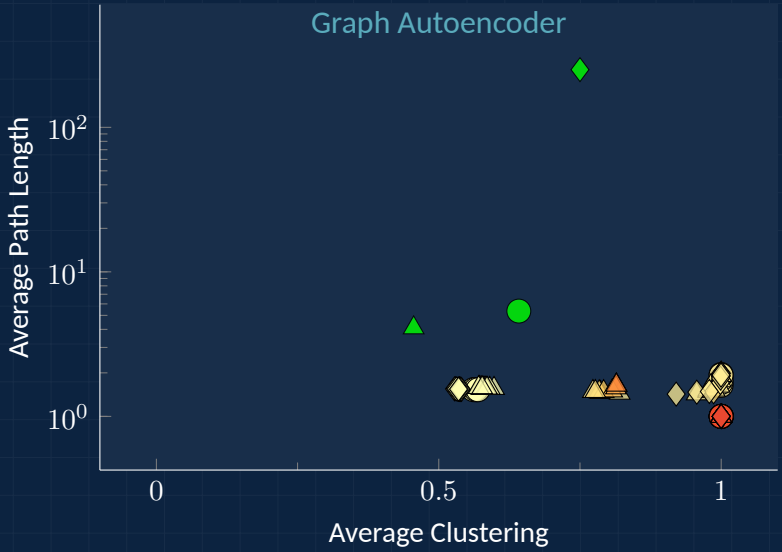
Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Infinity Mirror Test
Motivation
Methodology
Results
Conclusion
Wrapping Up

Tracking Evolution: Clustering & Geodesics



Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Infinity Mirror Test
Motivation
Methodology
Results
Conclusion
Wrapping Up

Tracking Evolution: Clustering & Geodesics



Scalable and Interpretable Graph Modeling with Graph Grammars
bit.ly/satyaki-slides
Introduction
Grammar Extraction
Graph Generation
Infinity Mirror Test
Motivation
Methodology
Results
Conclusion
Wrapping Up

Takeaways

Main Takeaways

- *Confirms* previously known biases in Kronecker graphs

Takeaways

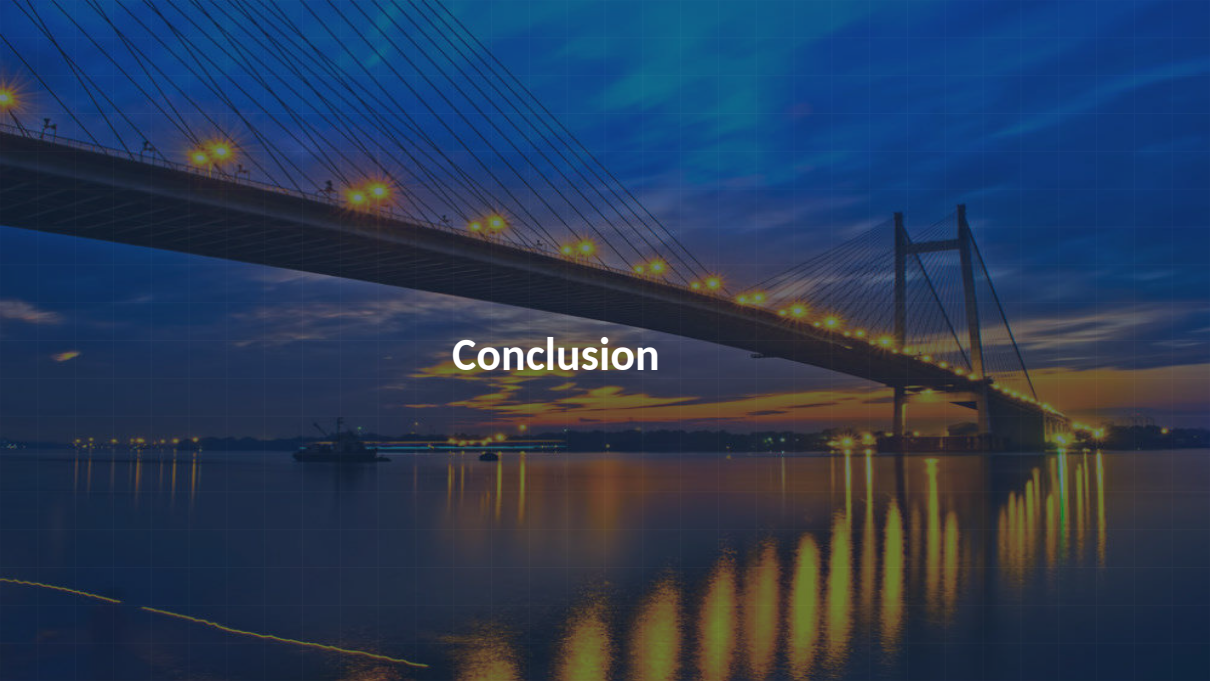
Main Takeaways

- *Confirms* previously known biases in Kronecker graphs
- Uncovers *unique* distortion patterns in popular graph models

Takeaways

Main Takeaways

- *Confirms* previously known biases in Kronecker graphs
- Uncovers *unique* distortion patterns in popular graph models
- A tool to design *better*, more parsimonious models



Conclusion

Summary

Vertex Replacement Grammars

- A new class of *scalable*, *interpretable* graph models
- Preserves both *topology* and *attribute* similarities

Summary

Vertex Replacement Grammars

- A new class of *scalable*, *interpretable* graph models
- Preserves both *topology* and *attribute* similarities

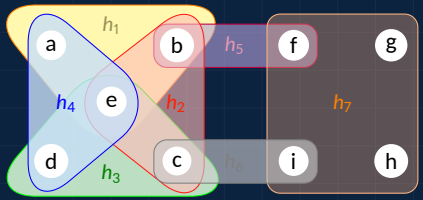
Infinity Mirror Test

- A *novel* stress-test for graph models
- Reveals new *biases* prompting further investigation

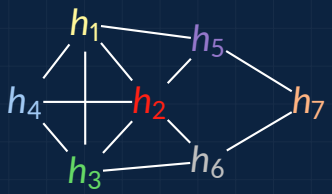
What's Next

Hypergraph Grammars

- New extraction method for HRGs *bypassing* tree decompositions
- Should improve both *stability* and scalability



Hypergraph *H*



Line Graph of *H*

Life at Indiana University

What's Next

Postdoctoral Fellow @ IU

- Work with *Santo Fortunato* on the MINERVA project
- Understanding the *process* of scientific discovery
- Studying the *dynamics* of citation and collaboration networks

Thank You!

Minimum Description Length Principle

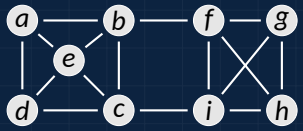
The Core Idea

- The best model is the one that leads to the best compression

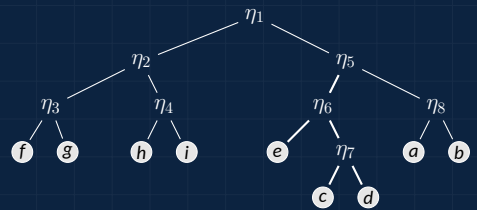
$$\mathcal{M}^* = \text{argmin} \{ \mathcal{L}(\mathcal{M}) + \mathcal{L}(\mathcal{D} \mid \mathcal{M}) \}$$

- $\mathcal{L}(\mathcal{M})$ is the length of the model
- $\mathcal{L}(\mathcal{D} \mid \mathcal{M})$ is the length of data compressed by the model

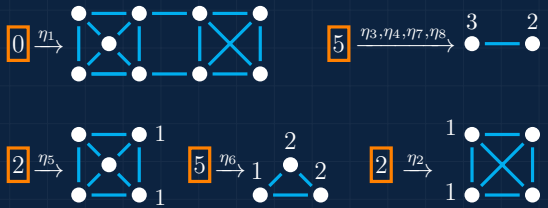
CNRG extraction and MDL



(a) Example graph H with 9 nodes and 16 edges



(b) An example dendrogram $\mathcal{D}$



(c) All possible rules that can be obtained from $\mathcal{D}$

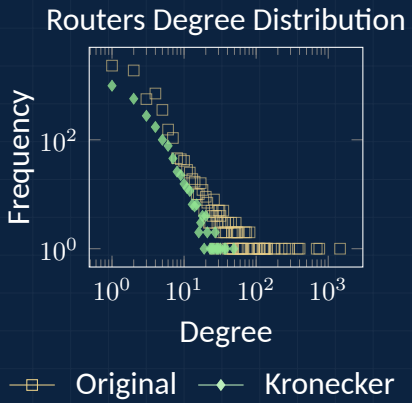
Comparing Graphs

Distribution Based

- Degree distribution
- PageRank centrality
- . . .

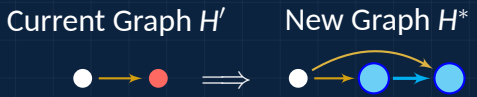
Topology Based

- Graph Edit Distance (GED)
- Graphlet Correlation Distance (GCD)
- DeltaCon - Personalized PR
- . . .



Kemp-Tenenbaum (KT) Graph Grammars

(A) A KT grammar rule



(B) Example of rule application

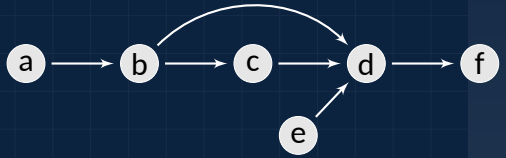
replaced new boundary

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

- Specify *max* rule size μ
- Find *connected* sets of up to size μ



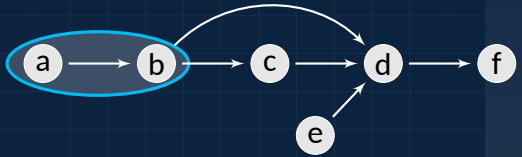
Enumerating all connected size μ subgraphs in *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

- Specify *max* rule size μ
- Find *connected* sets of up to size μ



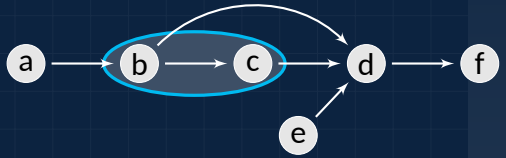
Enumerating all connected size μ subgraphs in *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

- Specify *max* rule size μ
- Find *connected* sets of up to size μ



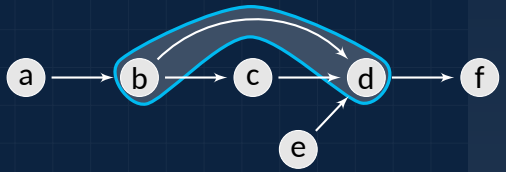
Enumerating all connected size μ subgraphs in *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

- Specify *max* rule size μ
- Find *connected* sets of up to size μ



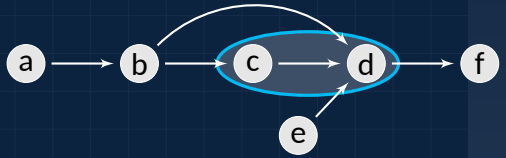
Enumerating all connected size μ subgraphs in *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

- Specify *max* rule size μ
- Find *connected* sets of up to size μ



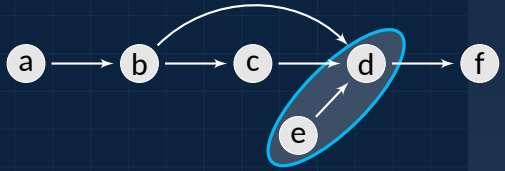
Enumerating all connected size μ subgraphs in *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

- Specify *max* rule size μ
- Find *connected* sets of up to size μ



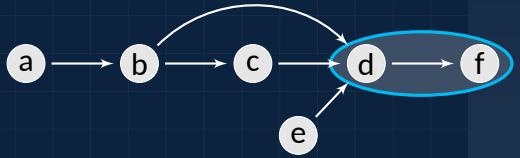
Enumerating all connected size μ subgraphs in *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

- Specify *max* rule size μ
- Find *connected* sets of up to size μ



Enumerating all connected size μ subgraphs in *H*

Bottom-up Graph Grammar Extraction (BUGGE)

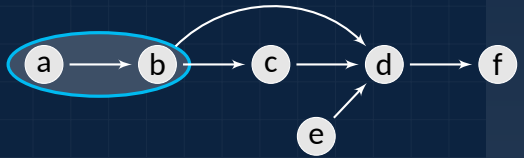
$(\mu = 2)$

(1) Enumerate Subgraphs

(2) Best Rule Selection

Use MDL principle to find the rule which *compresses* the graph the most

(A) Best rule *R*



(B) *R* appears 4 times in Graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

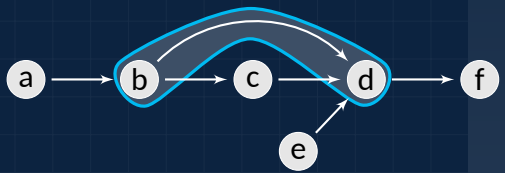
$(\mu = 2)$

(1) Enumerate Subgraphs

(2) Best Rule Selection

Use MDL principle to find the rule which *compresses* the graph the most

(A) Best rule *R*



(B) *R* appears 4 times in Graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

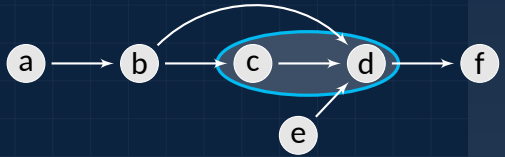
$(\mu = 2)$

(1) Enumerate Subgraphs

(2) Best Rule Selection

Use MDL principle to find the rule which *compresses* the graph the most

(A) Best rule *R*



(B) *R* appears 4 times in Graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

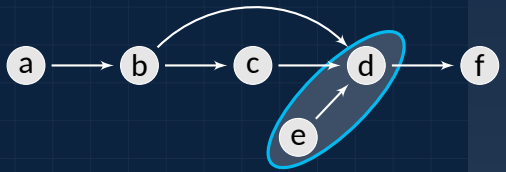
$(\mu = 2)$

(1) Enumerate Subgraphs

(2) Best Rule Selection

Use MDL principle to find the rule which *compresses* the graph the most

(A) Best rule *R*



(B) *R* appears 4 times in Graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

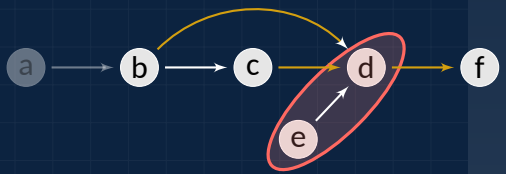
(1) Enumerate Subgraphs

(2) Best Rule Selection

(3) Extract Best Rule

- Match rule *RHS* in the graph
- *Collapse* matched nodes
- *Repeat* until the graph is empty

(A) Best rule *R*



(B) Current graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

(2) Best Rule Selection

(3) Extract Best Rule

- Match rule *RHS* in the graph
- *Collapse* matched nodes
- *Repeat* until the graph is empty

(A) Best rule *R*

(B) Current graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

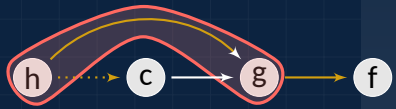
(1) Enumerate Subgraphs

(2) Best Rule Selection

(3) Extract Best Rule

- Match rule *RHS* in the graph
- *Collapse* matched nodes
- *Repeat* until the graph is empty

(A) Best rule *R*



(B) Current graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

(2) Best Rule Selection

(3) Extract Best Rule

- Match rule *RHS* in the graph
- *Collapse* matched nodes
- *Repeat* until the graph is empty

(A) Best rule *R*

(B) Current graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

(2) Best Rule Selection

(3) Extract Best Rule

- Match rule *RHS* in the graph
- *Collapse* matched nodes
- *Repeat* until the graph is empty

(A) Best rule *R*

(B) Current graph *H*

Bottom-up Graph Grammar Extraction (BUGGE)

$(\mu = 2)$

(1) Enumerate Subgraphs

(2) Best Rule Selection

(3) Extract Best Rule

- Match rule *RHS* in the graph
- *Collapse* matched nodes
- *Repeat* until the graph is empty

(A) Best rule *R*

(B) Current graph *H*

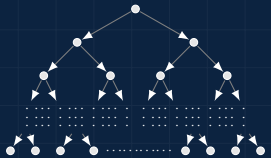
Results from Synthetic Graphs

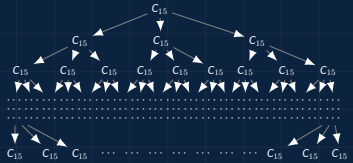
I: Balanced Binary Tree

II: Tree of Directed C_{15} Rings

III: Regular Ring Lattice

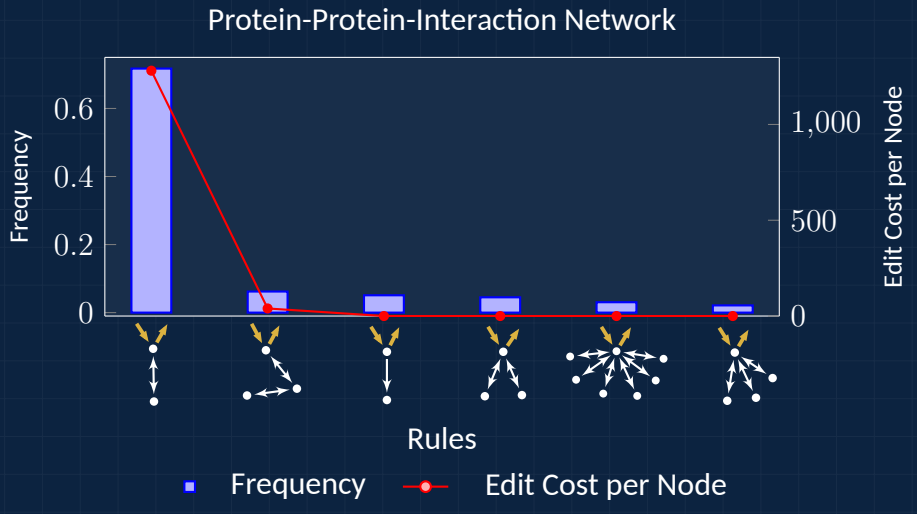
Synthetic Graphs





BUGGE Rules

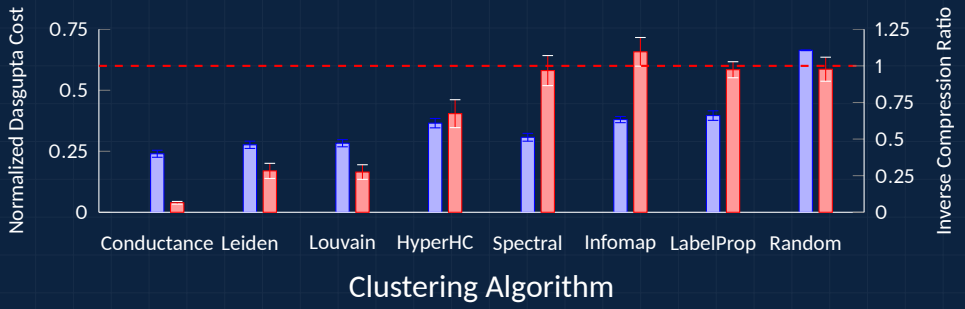
Most Frequent Rules Extracted from a PPI Network



CNRG Results

	EuCore		PolBlogs		OpenFlights	
	GCD	λ-dist	GCD	λ-dist	GCD	λ-dist
ChungLu	0.409	**0.803**	0.466	**1.234**	1.1116	**0.614**
HRG	0.229	8.091	1.196	4.407	1.2442	2.761
DC-SBM	**0.180**	2.057	0.262	4.186	0.8414	3.534
BTER	–	–	0.352	7.505	0.832	4.936
Kronecker	0.3164	11.802	1.302	14.31	1.83	10.459
CNRGE	0.233	4.969	**0.212**	4.276	**0.2832**	3.581

CNRG Results

	GrQc		PGP		Gnutella	
	GCD	λ-dist	GCD	λ-dist	GCD	λ-dist
ChungLu	2.657	**0.389**	2	**0.64**	1.02	0.42
HRG	1.99	4.41	–	–	2	5
DC-SBM	2.065	2.202	1.39	2.29	–	–
BTER	2.231	0.439	1.61	0.832	1.10	0.474
Kronecker	3.87	5.468	2.882	3.54	3.31	5.96
CNRGE	**1.067**	0.723	**0.448**	1.329	**0.41**	**0.20**

Dendrogram Selection