Modeling Graphs with Vertex Replacement Grammars

Satyaki Sikdar, Justus Hibshman, Tim Weninger

University of Notre Dame, IN, USA

November 10, 2019

bit.ly/graph_grammar

String Context Free Grammars

Production Rules

R_1 : S $\rightarrow$ NP VP
R_2 : NP $\rightarrow$ the N
R_3 : VP $\rightarrow$ V NP
R_4 : V $\rightarrow$ sings
R_5 : V $\rightarrow$ eats
R_6 : N $\rightarrow$ cat
R_7 : N $\rightarrow$ song
R_8 : N $\rightarrow$ canary

String Context Free Grammars

Production Rules

R_1 : S $\rightarrow$ NP VP
R_2 : NP $\rightarrow$ the N
R_3 : VP $\rightarrow$ V NP
R_4 : V $\rightarrow$ sings
R_5 : V $\rightarrow$ eats
R_6 : N $\rightarrow$ cat
R_7 : N $\rightarrow$ song
R_8 : N $\rightarrow$ canary

Deriving Strings

S $\overset{R_1}{\Longrightarrow}$ NP VP
$\overset{R_2}{\Longrightarrow}$ **the N** VP
$\overset{R_6}{\Longrightarrow}$ the **cat** VP
$\overset{R_3}{\Longrightarrow}$ the cat **V NP**
$\overset{R_4}{\Longrightarrow}$ the cat **sings** NP
$\overset{R_2}{\Longrightarrow}$ the cat sings **the N**
$\overset{R_7}{\Longrightarrow}$ the cat sings the **song**

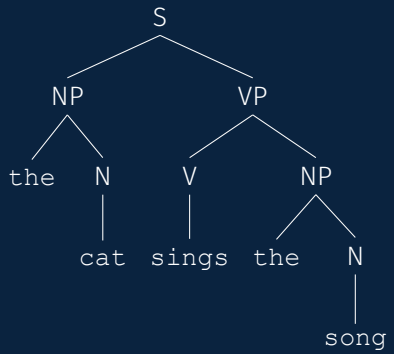
String Context Free Grammars

Production Rules

R_1 : S $\rightarrow$ NP VP
R_2 : NP $\rightarrow$ the N
R_3 : VP $\rightarrow$ V NP
R_4 : V $\rightarrow$ sings
R_5 : V $\rightarrow$ eats
R_6 : N $\rightarrow$ cat
R_7 : N $\rightarrow$ song
R_8 : N $\rightarrow$ canary

Deriving Strings

S $\overset{R_1}{\Rightarrow}$ NP VP
$\overset{R_2}{\Rightarrow}$ **the N** VP
$\overset{R_6}{\Rightarrow}$ the **cat** VP
$\overset{R_3}{\Rightarrow}$ the cat **V NP**
$\overset{R_4}{\Rightarrow}$ the cat **sings** NP
$\overset{R_2}{\Rightarrow}$ the cat sings **the N**
$\overset{R_7}{\Rightarrow}$ the cat sings the **song**



Parse Tree

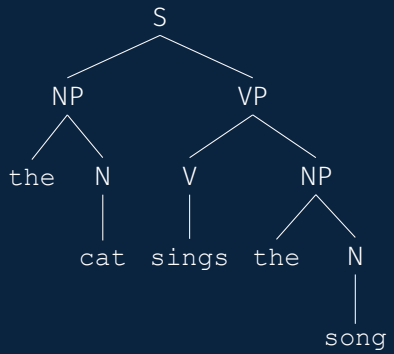
String Context Free Grammars

Production Rules

R_1 : S → NP VP
R_2 : NP → the N
R_3 : VP → V NP
R_4 : V → sings
R_5 : V → eats
R_6 : N → cat
R_7 : N → song
R_8 : N → canary

Deriving Strings

S $\overset{R_1}{\Rightarrow}$ NP VP
$\overset{R_2}{\Rightarrow}$ **the N** VP
$\overset{R_6}{\Rightarrow}$ the **cat** VP
$\overset{R_3}{\Rightarrow}$ the cat **V NP**
$\overset{R_4}{\Rightarrow}$ the cat **sings** NP
$\overset{R_2}{\Rightarrow}$ the cat sings **the N**
$\overset{R_7}{\Rightarrow}$ the cat sings the **song**



Parse Tree

Where did the production rules comes from? Can we learn them?

The Big Picture

Key Questions

The Big Picture

Key Questions

- Why is grammar important for language?

The Big Picture

Key Questions

- Why is grammar important for language?
- Can we do the same for graphs?

The Big Picture

Key Questions

- Why is grammar important for language?
- Can we do the same for graphs?

Goals

The Big Picture

Key Questions

- Why is grammar important for language?
- Can we do the same for graphs?

Goals

- Identifying *interesting* topological structures in a graph

The Big Picture

Key Questions

- Why is grammar important for language?
- Can we do the same for graphs?

Goals

- Identifying *interesting* topological structures in a graph
- Finding the building blocks via graph grammar rules

The Big Picture

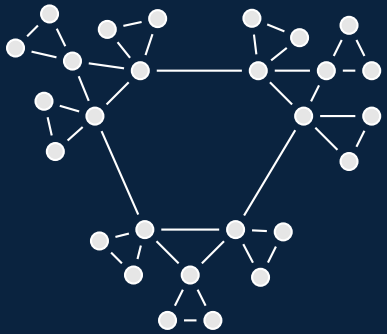
Key Questions

- Why is grammar important for language?
- Can we do the same for graphs?

Goals

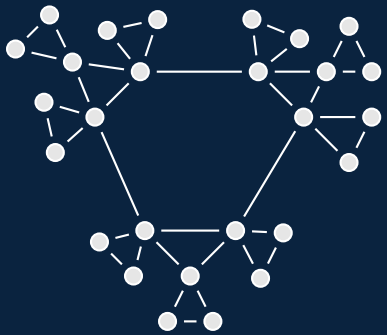
- Identifying *interesting* topological structures in a graph
- Finding the building blocks via graph grammar rules
- Using the building blocks to generate *similar* graphs

Intuition

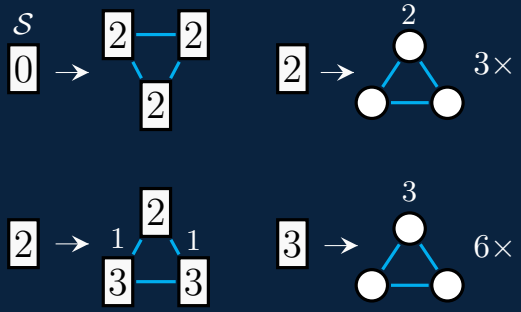


(a) Example Graph

Intuition

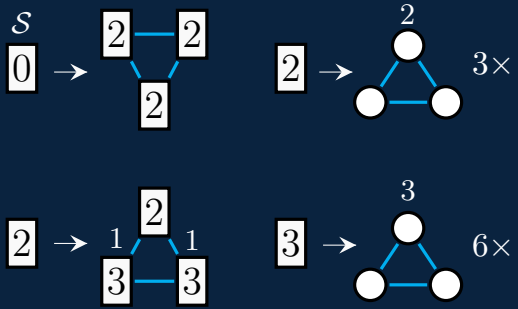


(a) Example Graph



(b) Example CNRG

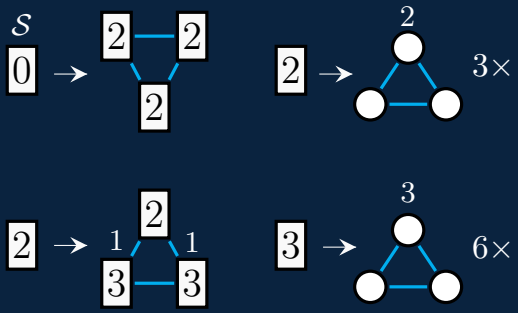
Formal Definition: Clustering-based Node Replacement Grammar (CNRG)



$G = \langle \Sigma, \Delta, \mathcal{P}, \mathcal{S} \rangle$

- Σ: alphabet of node labels
- Δ: alphabet of *terminal* node labels
- $\mathcal{P}$: set of production rules
- $\mathcal{S}$: starting graph

Formal Definition: Clustering-based Node Replacement Grammar (CNRG)



$G = \langle \Sigma, \Delta, \mathcal{P}, \mathcal{S} \rangle$

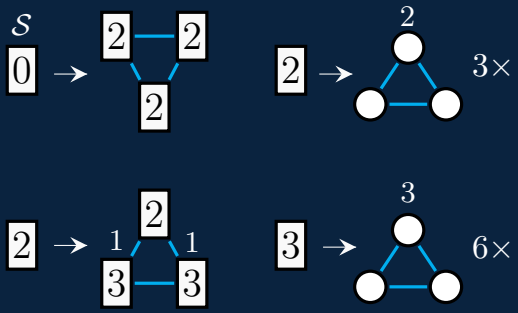
Σ: $\{$ ●, 2, 3 $\}$

Δ: $\{$ ● $\}$

$\mathcal{P}$: shown on the left

$\mathcal{S}$: 0

Formal Definition: Clustering-based Node Replacement Grammar (CNRG)



$G = \langle \Sigma, \Delta, \mathcal{P}, \mathcal{S} \rangle$

Σ: $\{$ ●, 2, 3 $\}$

Δ: $\{$ ● $\}$

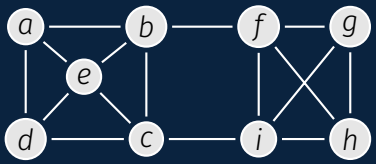
$\mathcal{P}$: shown on the left

$\mathcal{S}$: 0

Where did the production rules comes from? Can we learn them?

Extracting a CNRG: I

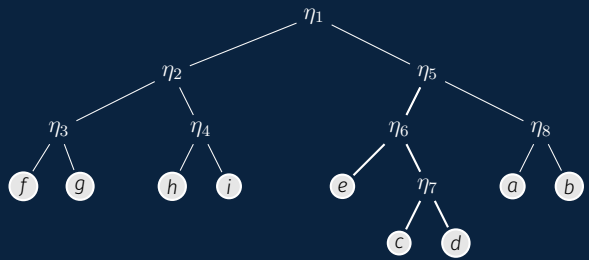
$(\mu = 4)$



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

Leiden
Louvain

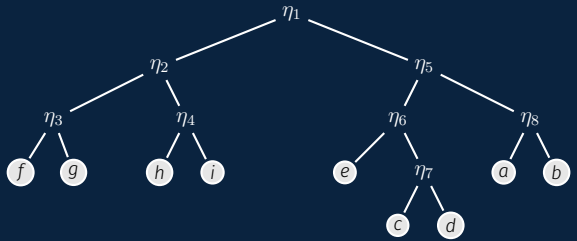
Spectral
MinCut



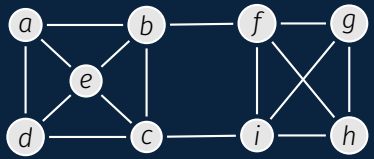
(b) An example dendrogram $\mathcal{D}$

Extracting a CNRG: II

$(\mu = 4)$



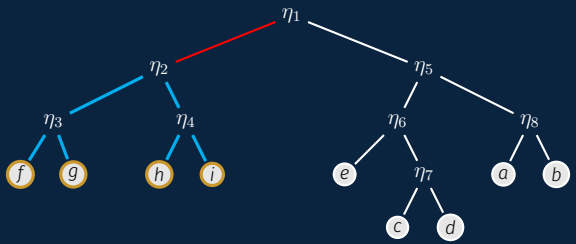
(a) Current Dendrogram $\mathcal{D}$



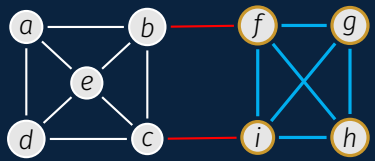
(b) Current Graph H

Extracting a CNRG: II

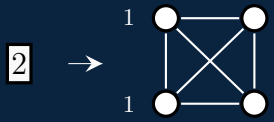
$(\mu = 4)$



(a) Current Dendrogram $\mathcal{D}$



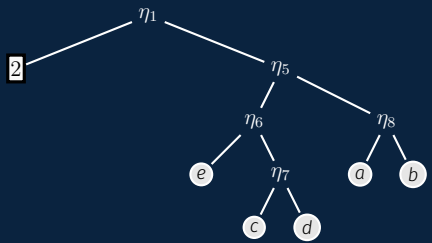
(b) Current Graph H



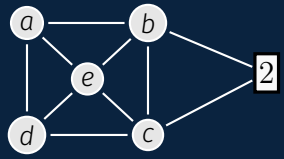
(c) Extracted CNRG Rule

Extracting a CNRG: II

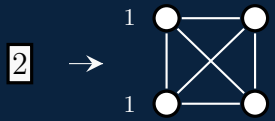
$(\mu = 4)$



(a) Updated Dendrogram $\mathcal{D}'$



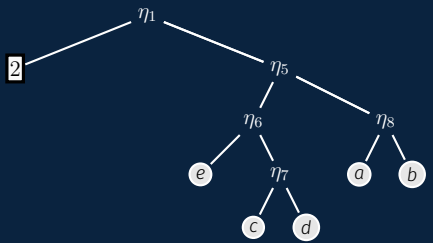
(b) Updated Graph H'



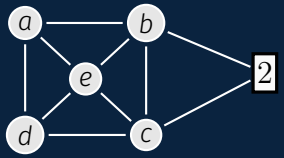
(c) Extracted CNRG Rule

Extracting a CNRG: III

$(\mu = 4)$



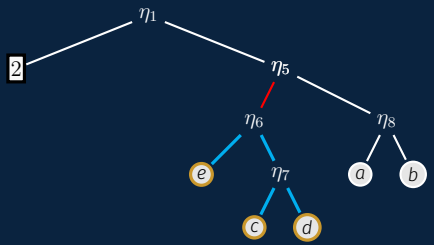
(a) Current Dendrogram $\mathcal{D}$



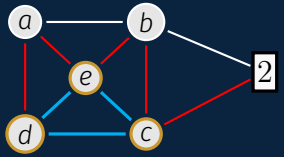
(b) Current Graph H

Extracting a CNRG: III

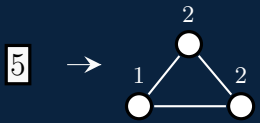
$(\mu = 4)$



(a) Current Dendrogram $\mathcal{D}$



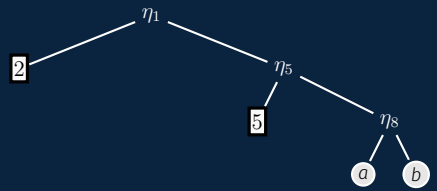
(b) Current Graph H



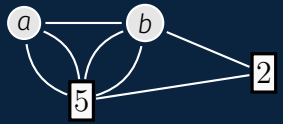
(c) Extracted CNRG Rule

Extracting a CNRG: III

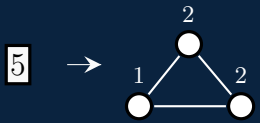
$(\mu = 4)$



(a) Updated Dendrogram $\mathcal{D}'$



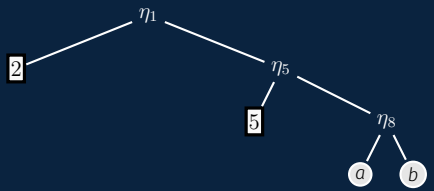
(b) Updated Graph H'



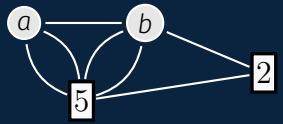
(c) Extracted CNRG Rule

Extracting a CNRG: IV

$(\mu = 4)$



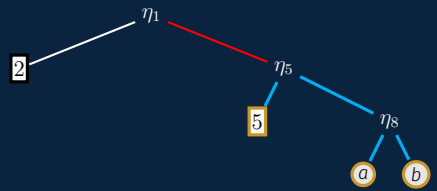
(a) Current Dendrogram $\mathcal{D}$



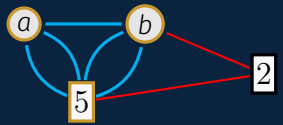
(b) Current Graph H

Extracting a CNRG: IV

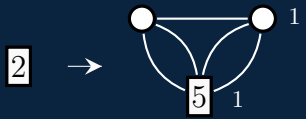
$(\mu = 4)$



(a) Current Dendrogram $\mathcal{D}$



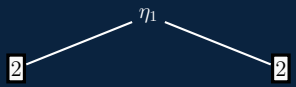
(b) Current Graph H



(c) Extracted CNRG Rule

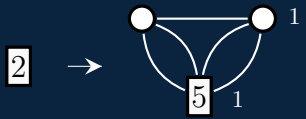
Extracting a CNRG: IV

$(\mu = 4)$



(a) Updated Dendrogram $\mathcal{D}'$

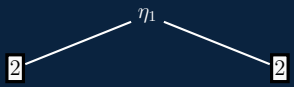
(b) Updated Graph H'



(c) Extracted CNRG Rule

Extracting a CNRG: V

$(\mu = 4)$

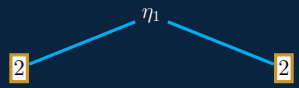


(a) Current Dendrogram $\mathcal{D}$

(b) Current Graph H

Extracting a CNRG: V

$(\mu = 4)$



(a) Current Dendrogram $\mathcal{D}$

(b) Current Graph H

(c) Extracted CNRG Rule

Extracting a CNRG: V

$(\mu = 4)$

(a) Updated Dendrogram $\mathcal{D}'$

(b) Updated Graph H'

(c) Extracted CNRG Rule

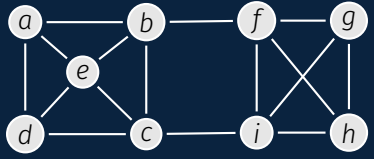
Extracting a CNRG: V

$(\mu = 4)$

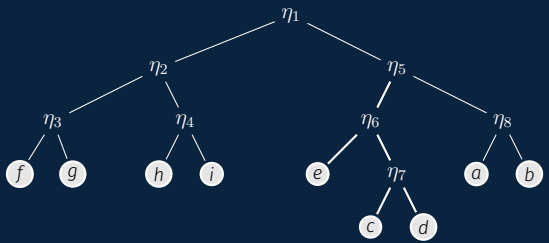
(a) Updated Dendrogram $\mathcal{T}$ **(b)** Updated Graph H'

(c) Extracted CNRG Rule

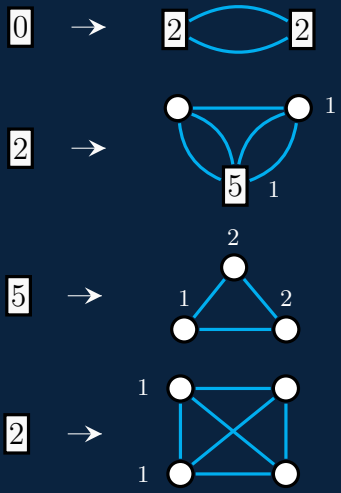
Extracted CNRG Rules



(a) Graph H

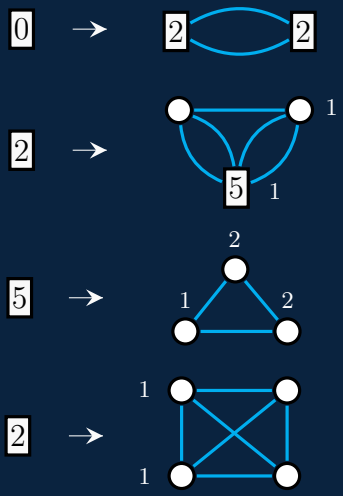


(b) Dendrogram $\mathcal{D}$



(c) Extracted CNRG Rules

Generating Graphs from a CNRG: I



Extracted CNRG Rules

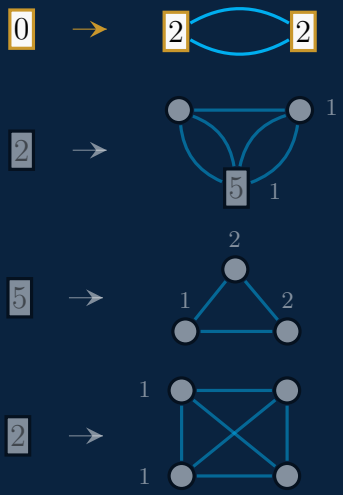
$\mathcal{S}$

0

Current Graph H'

Next Graph $\hat{H}$

Generating Graphs from a CNRG: I



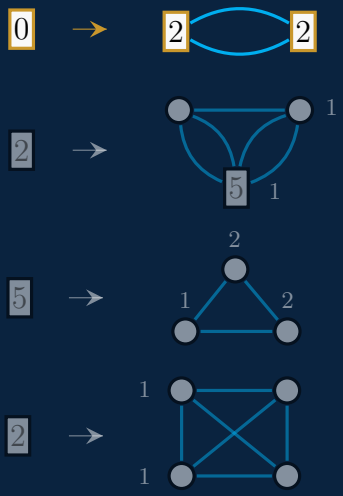
Extracted CNRG Rules

0

Current Graph H'

Next Graph $\hat{H}$

Generating Graphs from a CNRG: I

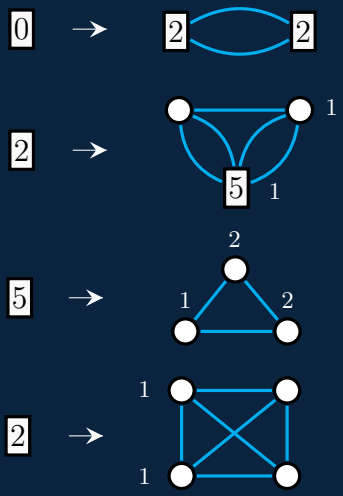


Extracted CNRG Rules

Current Graph H'

Next Graph $\hat{H}$

Generating Graphs from a CNRG: II

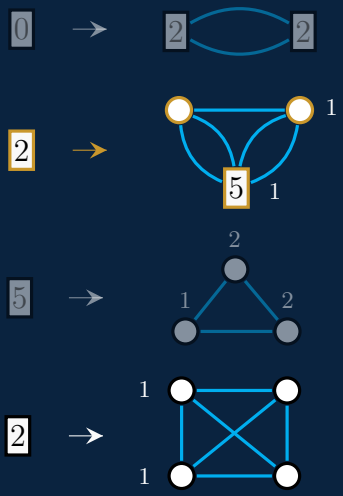


Extracted CNRG Rules

Current Graph H'

Next Graph $\hat{H}$

Generating Graphs from a CNRG: II

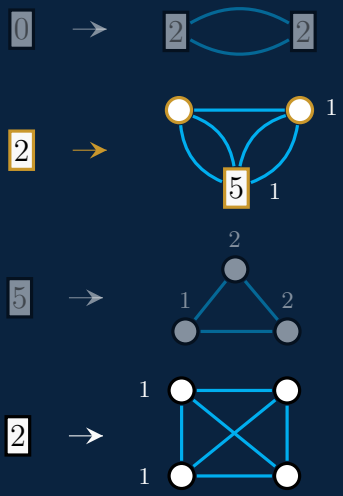


Extracted CNRG Rules

Current Graph H'

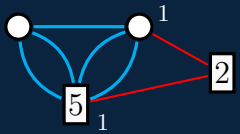
Next Graph $\hat{H}$

Generating Graphs from a CNRG: II



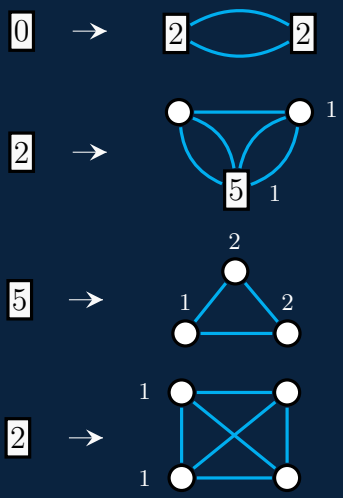
Extracted CNRG Rules

Current Graph H'

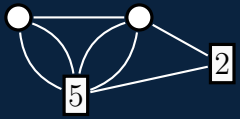


Next Graph $\hat{H}$

Generating Graphs from a CNRG: III



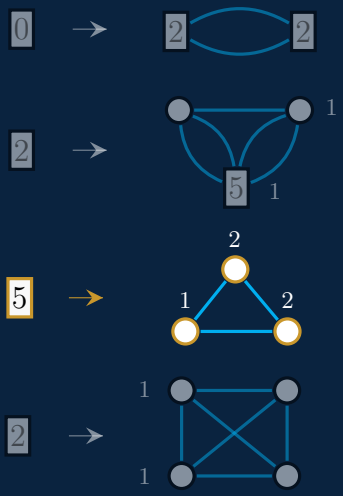
Extracted CNRG Rules



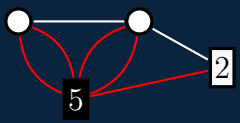
Current Graph H'

Next Graph $\hat{H}$

Generating Graphs from a CNRG: III



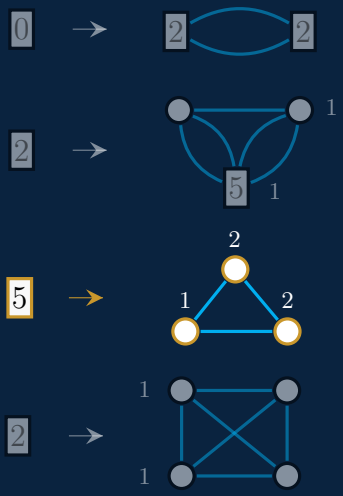
Extracted CNRG Rules



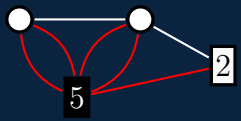
Current Graph H'

Next Graph $\hat{H}$

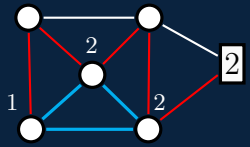
Generating Graphs from a CNRG: III



Extracted CNRG Rules

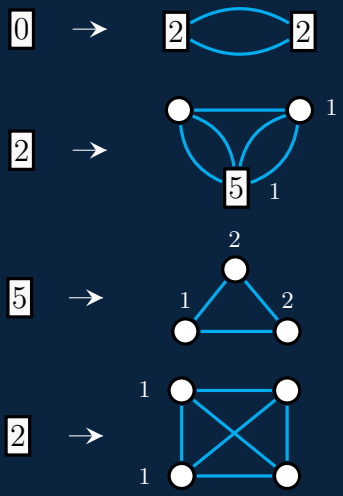


Current Graph H'

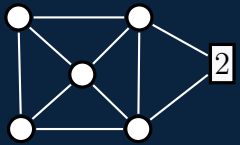


Next Graph $\hat{H}$

Generating Graphs from a CNRG: IV



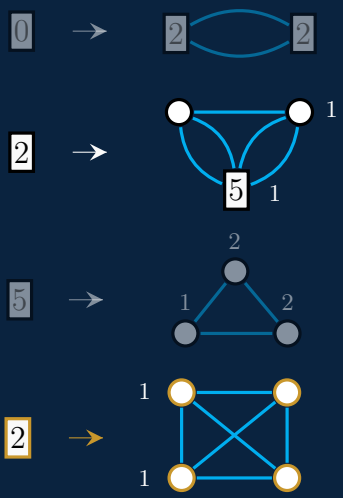
Extracted CNRG Rules



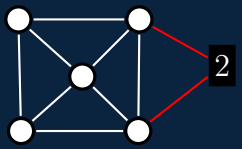
Current Graph H'

Next Graph $\hat{H}$

Generating Graphs from a CNRG: IV



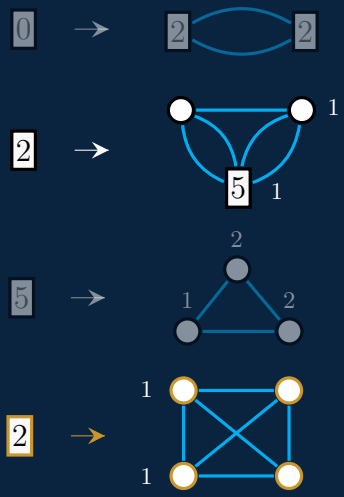
Extracted CNRG Rules



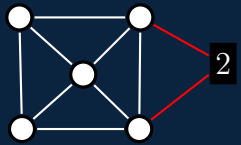
Current Graph H'

Next Graph $\hat{H}$

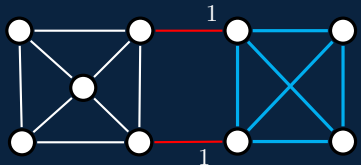
Generating Graphs from a CNRG: IV



Extracted CNRG Rules

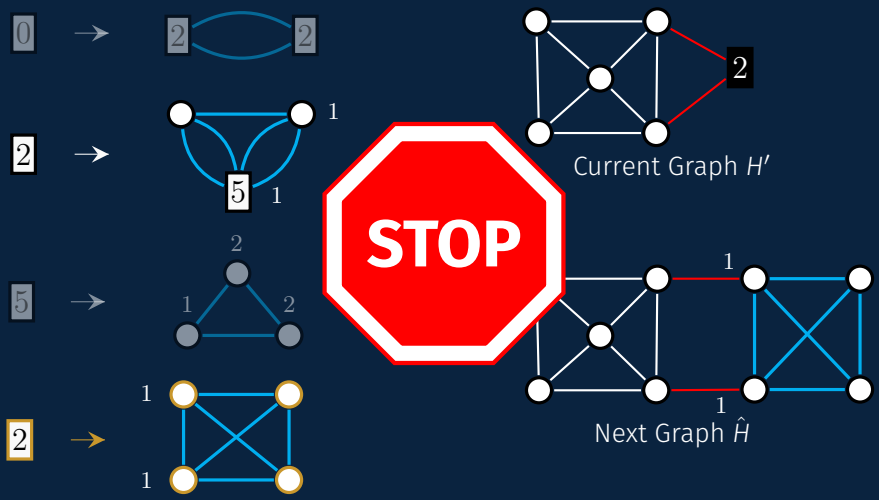


Current Graph H'

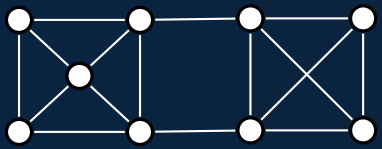


Next Graph $\hat{H}$

Generating Graphs from a CNRG: IV

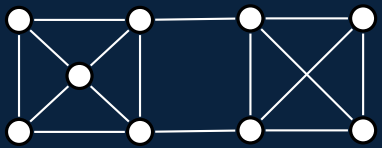


Generating Graphs from a CNRG

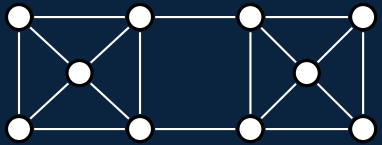
$\boxed{0} \stackrel{*}{\Rightarrow}$ 

Isomorphic with $p = \frac{1}{2}$!

Generating Graphs from a CNRG

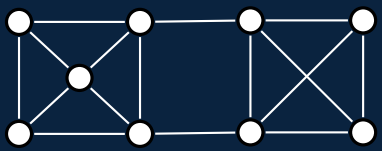
$\boxed{0} \overset{*}{\Rightarrow}$ 

Isomorphic with $p = \frac{1}{2}$!

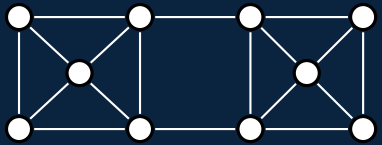
$\boxed{0} \overset{*}{\Rightarrow}$ 

$p = \frac{1}{4}$

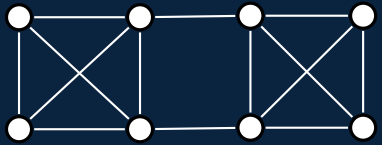
Generating Graphs from a CNRG

$\boxed{0} \Rightarrow^{*}$ 

Isomorphic with $p = \frac{1}{2}$!

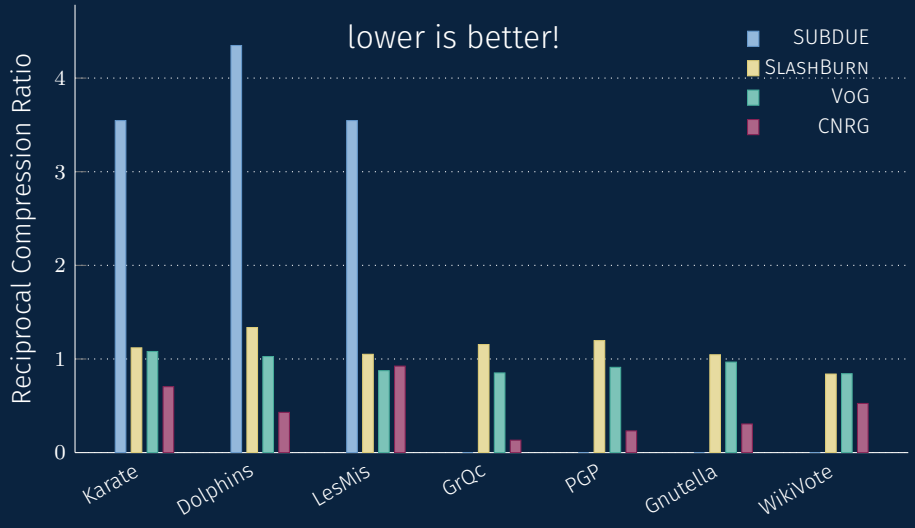
$\boxed{0} \Rightarrow^{*}$ 

$p = \frac{1}{4}$

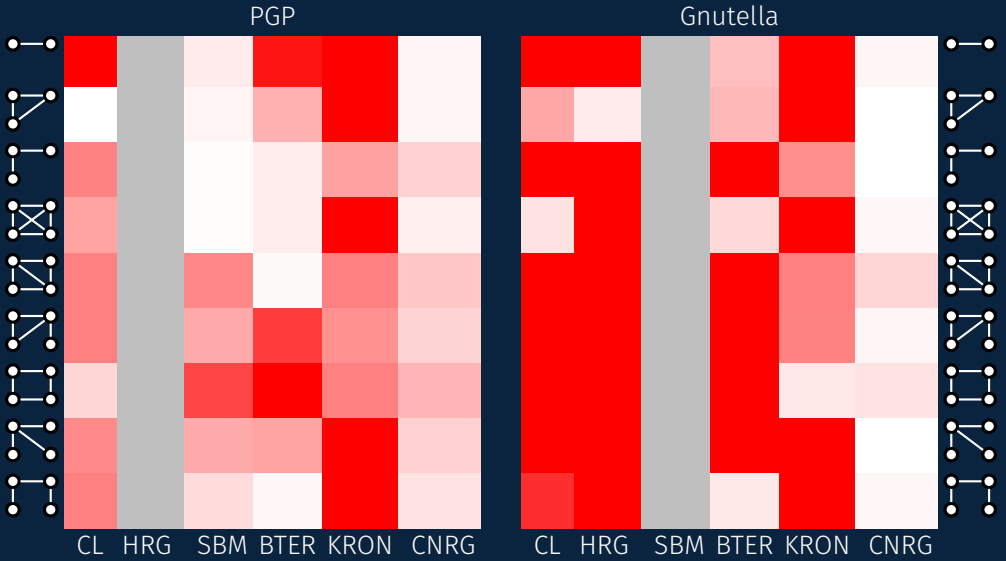
$\boxed{0} \Rightarrow^{*}$ 

$p = \frac{1}{4}$

Model Size Comparison



Graph Generation Quality Comparison (Lighter Red is Better)



Takeaways and Future Work

Main Takeaways

Takeaways and Future Work

Main Takeaways

- *Simple* and *compact* formalism borrowed from formal language theory

Takeaways and Future Work

Main Takeaways

- *Simple* and *compact* formalism borrowed from formal language theory
- *Scalable* to medium-large sized graphs

Takeaways and Future Work

Main Takeaways

- *Simple* and *compact* formalism borrowed from formal language theory
- *Scalable* to medium-large sized graphs
- *Faithful* graph generation

Takeaways and Future Work

Main Takeaways

- *Simple* and *compact* formalism borrowed from formal language theory
- *Scalable* to medium-large sized graphs
- *Faithful* graph generation

Key Questions

Takeaways and Future Work

Main Takeaways

- *Simple* and *compact* formalism borrowed from formal language theory
- *Scalable* to medium-large sized graphs
- *Faithful* graph generation

Key Questions

- Can we extract more meaningful rules?

Takeaways and Future Work

Main Takeaways

- *Simple* and *compact* formalism borrowed from formal language theory
- *Scalable* to medium-large sized graphs
- *Faithful* graph generation

Key Questions

- Can we extract more meaningful rules?
- Can we use CNRG as a null model for graphs?

Takeaways and Future Work

Main Takeaways

- *Simple* and *compact* formalism borrowed from formal language theory
- *Scalable* to medium-large sized graphs
- *Faithful* graph generation

Key Questions

- Can we extract more meaningful rules?
- Can we use CNRG as a null model for graphs?
- Can we adapt the formalism for attributed graphs?

Takeaways and Future Work

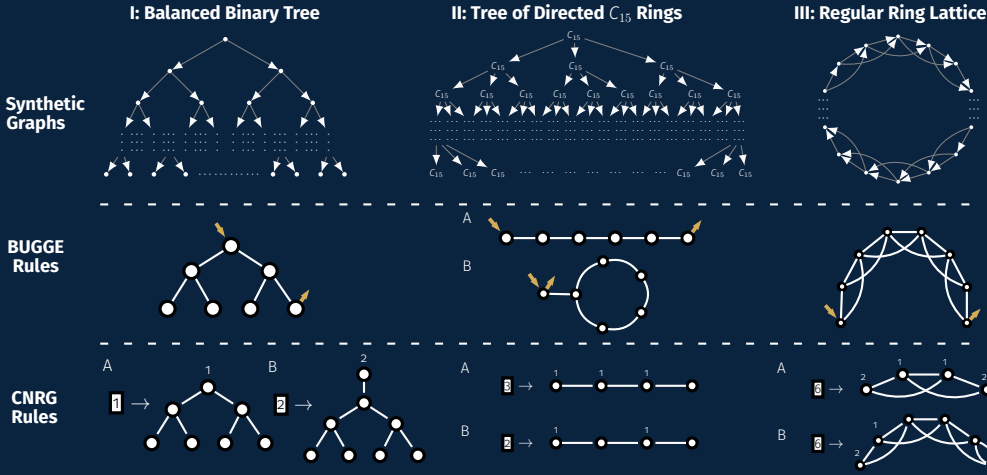
Main Takeaways

- *Simple* and *compact* formalism borrowed from formal language theory
- *Scalable* to medium-large sized graphs
- *Faithful* graph generation

Key Questions

- Can we extract more meaningful rules?
- Can we use CNRG as a null model for graphs?
- Can we adapt the formalism for attributed graphs?
- Can we utilize a related formalism to study and track changes in dynamic graphs?

What's Next? Making Sense of Grammar Rules



Towards Interpretable Graph Modeling with Vertex Replacement Grammars, J. Hibshman, S. Sikdar, and T. Weninger, accepted at IEEE BigData 2019.

Thanks!

ssikdar@nd.edu
@satyaki30