

Fine-Grained Complexity

Week 12

Input:

Genome (1D)

Today:

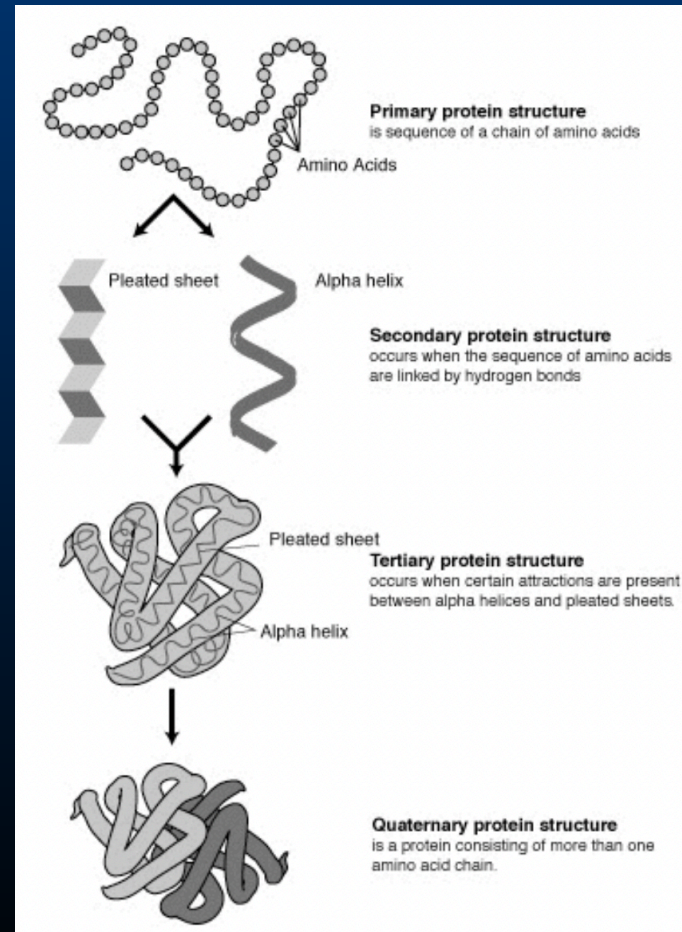
RNA Folding (2D)

Real Goal:

Protein Folding (3D)

NP-Hard

“Solved” by AlphaFold



From Wikipedia

RNA Folding

“Maximum matching on a line”

Input: A sequence in $\{A,C,G,T\}^n$.

Output: Maximum number of non-crossing matches.

A C C G A A C T C C G G T A

The diagram shows the RNA sequence ACCGA AACTCCGGTA. Red arcs connect the following pairs of bases: (4th A, 10th T), (5th A, 11th T), (6th C, 12th G), (7th C, 13th G), and (14th G, 15th T). These arcs do not cross each other.

Matches:

A - T

C - G

RNA Folding

“Maximum matching on a line”

Input: A sequence in $\{A,C,G,T\}^n$.

Output: Maximum number of non-crossing matches.



Matches:

A - T

C - G

RNA Folding

“Maximum matching on a line”

Input: A sequence in $\{A,C,G,T\}^n$.

Output: Maximum number of non-crossing matches.



Matches:

A - T

C - G

In general: A sequence in $(\Sigma \cup \Sigma')^n$.

$\forall \sigma \in \Sigma$: there is a unique match $\sigma' \in \Sigma'$.

List of RNA Folding Algorithms

Single sequence secondary structure prediction. [[edit](#)]

Name	Description	Knots <small>[Note 1]</small>	Links	References
CentroidFold	Secondary structure prediction based on generalized centroid estimator	No	sourcecode@webserv@	[1]
CentroidHomfold	Secondary structure prediction by using homologous sequence information	No	sourcecode@webserv@	[2]
Context Fold	An RNA secondary structure prediction software based on feature-rich trained scoring models.	No	sourcecode@webserv@	[3]
CONTRAFold	Secondary structure prediction method based on conditional log-linear models (CLLMs), a flexible class of probabilistic models which generalize upon SCFGs by using discriminative training and feature-rich scoring.	No	sourcecode@webserv@	[4]
CytoFold	Secondary structure prediction method based on placement of helices allowing complex pseudoknots.	Yes	webserv@	[5]
GTFold	Fast and scalable multicore code for predicting RNA secondary structure.	No	link@sourcecode@	[6]
IPknot	Fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming.	Yes	sourcecode@webserv@	[7]
KineFold	Folding kinetics of RNA sequences including pseudoknots by including an implementation of the partition function for knots.	Yes	linuxbinary@webserv@	[8][9]
Mfold	MFE (Minimum Free Energy) RNA structure prediction algorithm.	No	sourcecode@webserv@	[10]
Pknots	A dynamic programming algorithm for optimal RNA pseudoknot prediction using the nearest neighbour energy model.	Yes	sourcecode@	[11]
PknotsRG	A dynamic programming algorithm for the prediction of a restricted class (H-type) of RNA pseudoknots.	Yes	sourcecode@webserv@	[12]
pKiss	A dynamic programming algorithm for the prediction of a restricted class (H-type and kissing hairpins) of RNA pseudoknots.	Yes	sourcecode@webserv@	[13]
RNA123	Secondary structure prediction via thermodynamic-based folding algorithms and novel structure-based sequence alignment specific for RNA.	Yes	webserv@	
RNAfold	MFE RNA structure prediction algorithm. Includes an implementation of the partition function for computing basepair probabilities and circular RNA folding.	No	sourcecode@webserv@	[10][14][15][16][17]
RNAshapes	MFE RNA structure prediction based on abstract shapes. Shape abstraction retains adjacency and nesting of structural features, but disregards helix lengths, thus reduces the number of suboptimal solutions without losing significant information. Furthermore, shapes represent classes of structures for which probabilities based on Boltzmann-weighted energies can be computed.	No	source & binaries@webserv@	[18][19]
RNAstructure	A program to predict lowest free energy structures and base pair probabilities for RNA or DNA sequences. Programs are also available to predict <i>maximum expected accuracy</i> structures and these can include pseudoknots. Structure prediction can be constrained using experimental data, including SHAPE, enzymatic cleavage, and chemical modification accessibility. Graphical user interfaces are available for Windows, Mac OS X, Linux. Programs are also available for use with Unix-style text interfaces. Also, a C++ class library is available.	Yes	source & binaries@webserv@	[20][21]
SARNA-Predict	RNA Secondary structure prediction method based on simulated annealing. It can also predict structure with pseudoknots.	Yes	link@	[22]
vsfold/vs subopt	Folds and predicts RNA secondary structure and pseudoknots using an entropy model derived from polymer physics. The program vs_subopt computes suboptimal structures based on the free energy landscape derived from vsfold5.	Yes	webserv@	[23][24]
Sfold	Statistical sampling of all possible structures. The sampling is weighted by partition function probabilities.	No	webserv@	[25][26][27][28]
UNAFold	The UNAFold software package is an integrated collection of programs that simulate folding, hybridization, and melting pathways for one or two single-stranded nucleic acid sequences.	No	sourcecode@	[29]
CrumpIe	Simple, cleanly written software to produce the full set of possible secondary structures for one sequence, given optional constraints.	No	sourcecode@	[30]
Sliding Windows & Assembly	Sliding windows and assembly is a tool chain for folding long series of similar hairpins.	No	sourcecode@	[30]
SwiSpot	Command-line utility for predicting alternative (secondary) configurations of riboswitches. It is based on the prediction of the so-called switching sequence, to subsequently constrain the folding of the two functional structures.	No	sourcecode@	[31]

Special pages
 Permanent link
 Page information
 Wikidata item
 Cite this page

Print/export
 Create a book
 Download as PDF
 Printable version

Languages +
 Add links

RNA Folding is in P

Input: A sequence in $\{A,C,G,T\}^n$.

Output: Maximum number of non-crossing matches.

A C C G A A C T C C A C C G G T A

Matches:

A - T

C - G

$$D[i, j] = OPT(x[i, \dots, j]) = \max \left\{ \right.$$

RNA Folding is in P

Input: A sequence in $\{A,C,G,T\}^n$.

Output: Maximum number of non-crossing matches.



Matches:

A - T

C - G

$$D[i, j] = OPT(x[i, \dots, j]) = \max \left\{ \begin{array}{l} (x[i] = x[j]) + D[i + 1, j - 1] \\ \dots \end{array} \right.$$

RNA Folding is in P

Input: A sequence in $\{A,C,G,T\}^n$.

Output: Maximum number of non-crossing matches.



Matches:

A - T

C - G

$$D[i, j] = OPT(x[i, \dots, j]) = \max \left\{ \begin{array}{l} (x[i] = x[j]) + D[i+1, j-1] \\ \max_{i \leq k < j} D[i, k] + D[k+1, j] \end{array} \right.$$

$$n^2 \cdot n$$

RNA Folding is in P

Input: A sequence in $\{A, C, G, T\}^n$.

Output: Maximum number of non-crossing matches.



Matches:

A - T

C - G

$$D[i, j] = OPT(x[i, \dots, j]) = \max \begin{cases} (x[i] = x[j]) + D[i + 1, j - 1] \\ \max_{i \leq k < j} D[i, k] + D[k + 1, j] \end{cases}$$

Thm: RNA Folding is in $O(n^3)$ time.

Quadratic Lower Bound?

Quadratic Lower Bound?

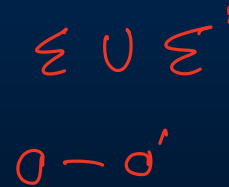
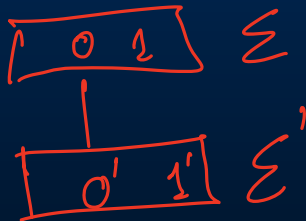
Longest Common Subsequence (LCS)



RNA Folding

n

$2n$



Thm: $O(n^{2-\epsilon})$ for RNA Folding refutes SETH.

Quadratic Lower Bound?

Longest Common Subsequence (LCS)



RNA Folding

x

y

$w =$

$$OPT_{RNA}(w) = LCS(x, y)$$

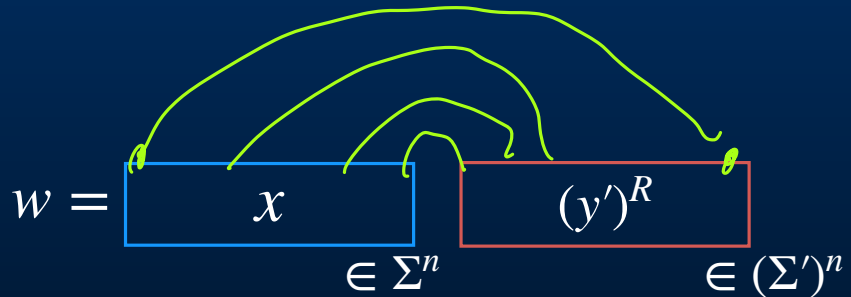
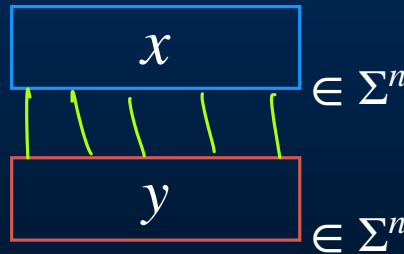
Thm: $O(n^{2-\epsilon})$ for RNA Folding refutes SETH.

Quadratic Lower Bound?

Longest Common Subsequence (LCS)



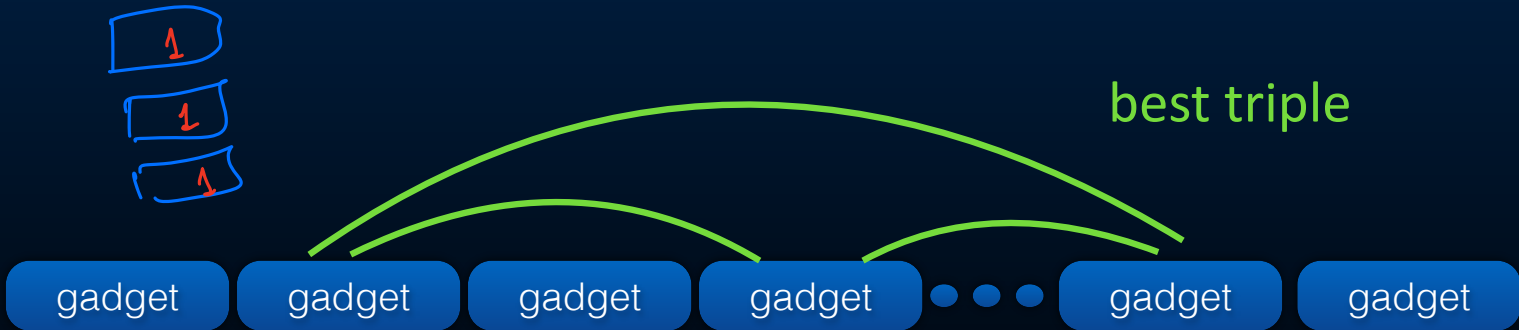
RNA Folding



$$OPT_{RNA}(w) = LCS(x, y)$$

Thm: $O(n^{2-\epsilon})$ for RNA Folding refutes SETH.

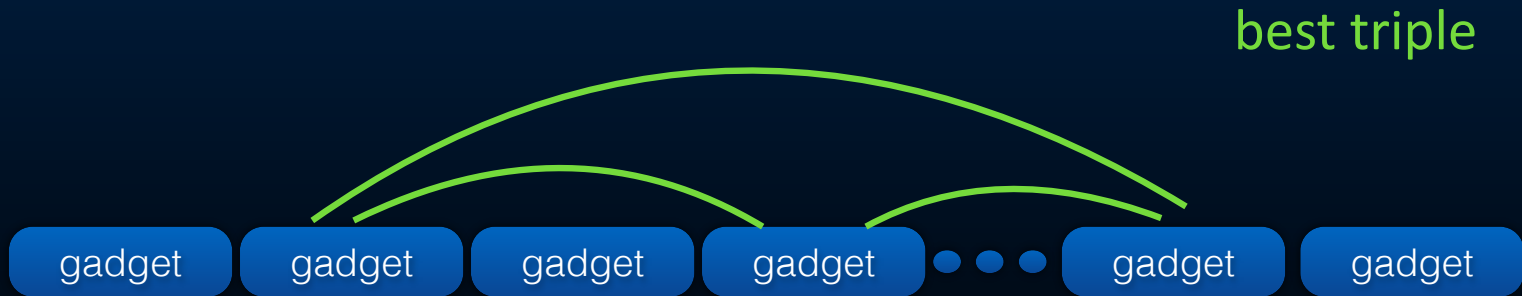
Higher Lower Bounds?



Higher Lower Bounds?

$\Omega(n^\omega)?$

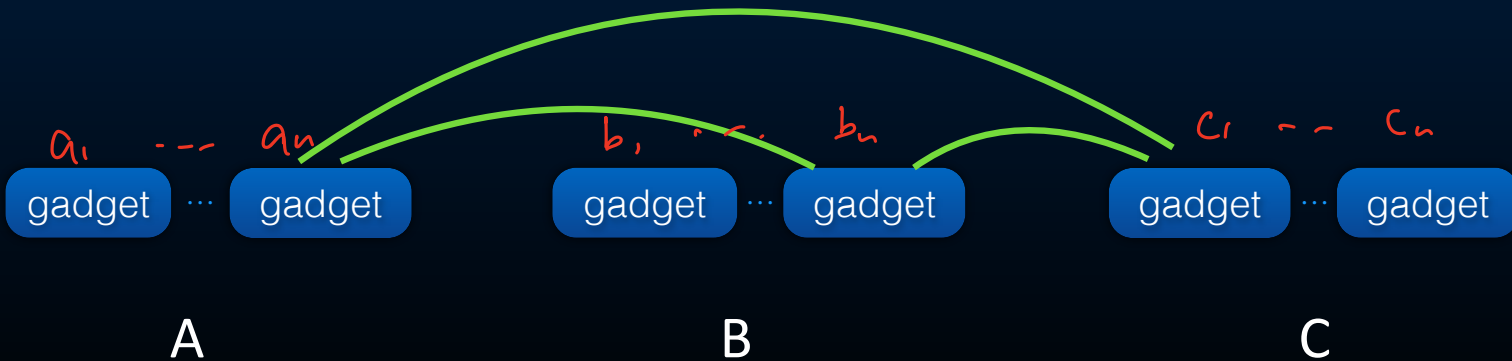
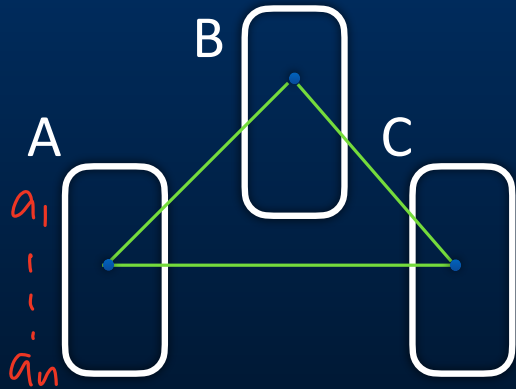
Attempt: Reduce from Triangle detection.



Higher Lower Bounds?

$\Omega(n^\omega)?$

Attempt: Reduce from Triangle detection.



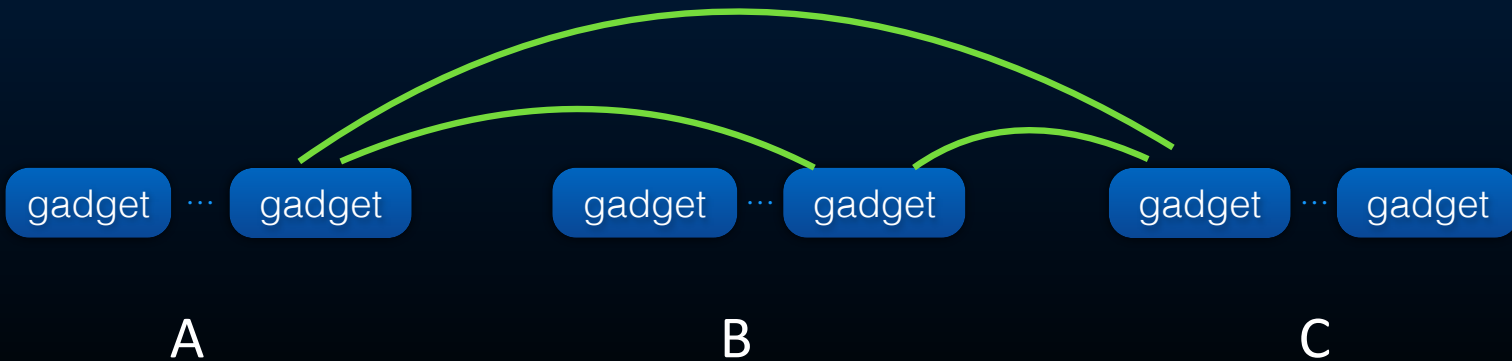
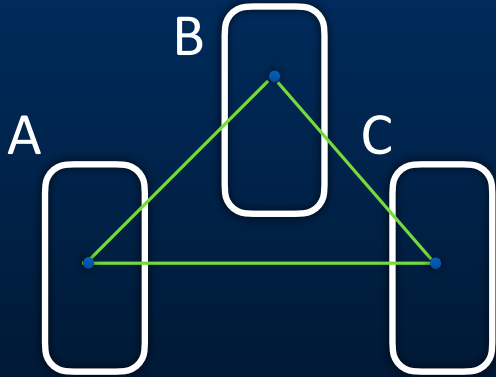
Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that:

$$OPT(\overset{\xi}{a} \quad \overset{\xi'}{b}) > \tau$$

iff

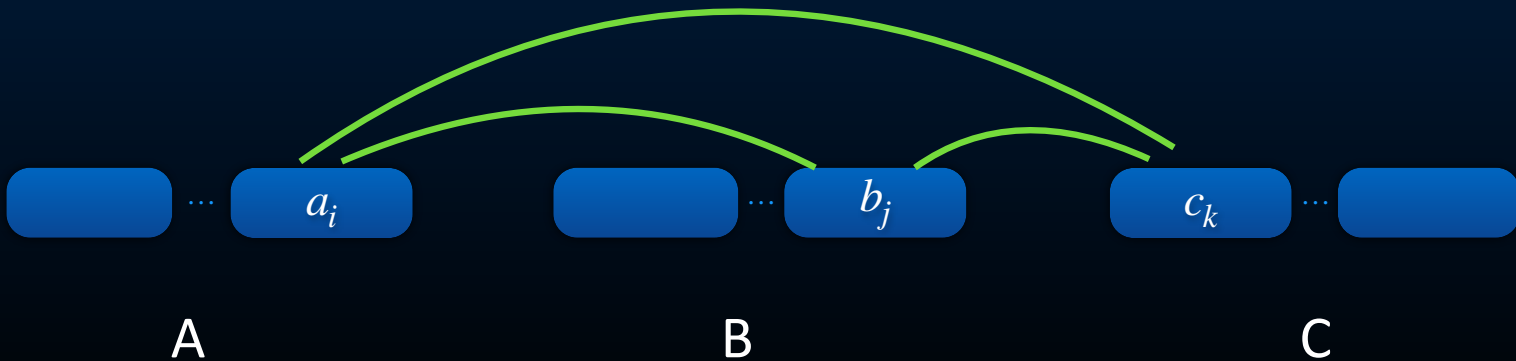
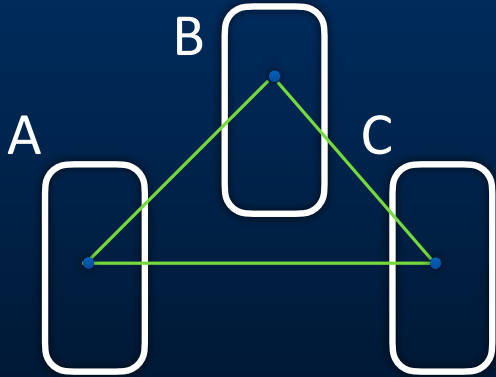
$$(a, b) \in E$$



Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that: $OPT(\text{blue } a, \text{red } b) > \tau$

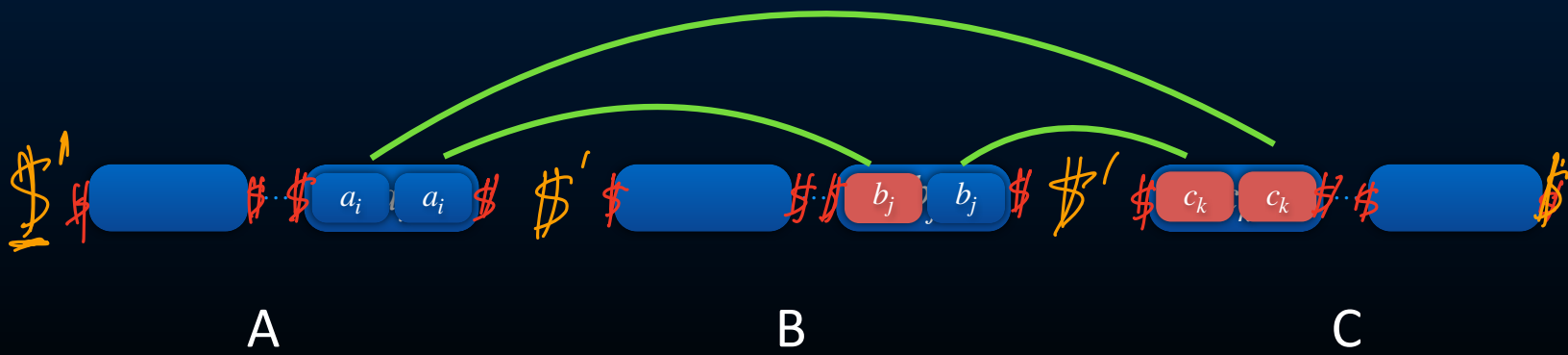
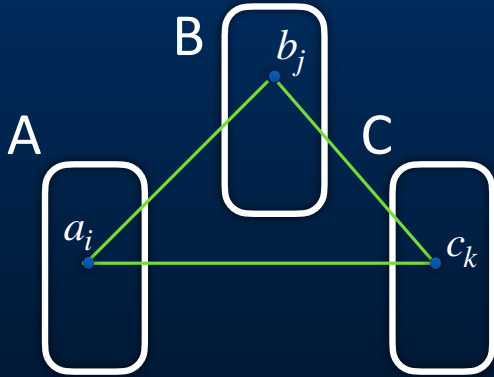
iff
 $(a, b) \in E$



Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that: $OPT(\text{blue } a, \text{red } b) > \tau$

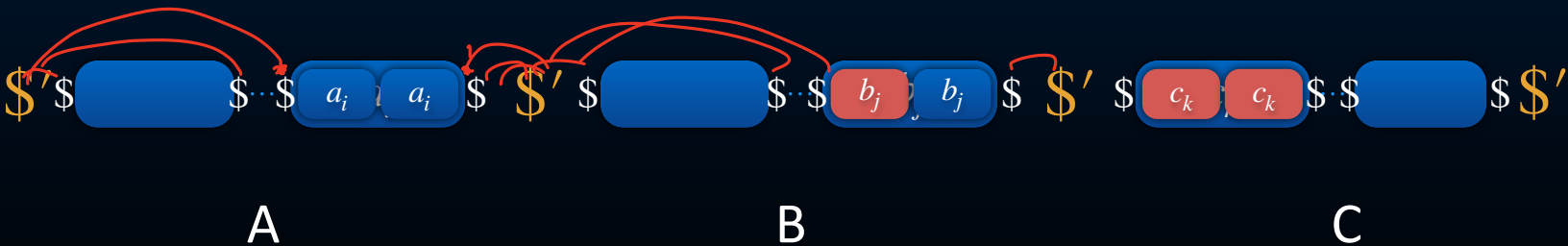
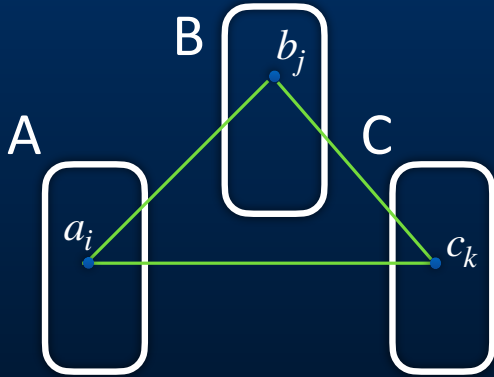
iff
 $(a, b) \in E$



Higher Lower Bounds? $\Omega(n^\omega)$?

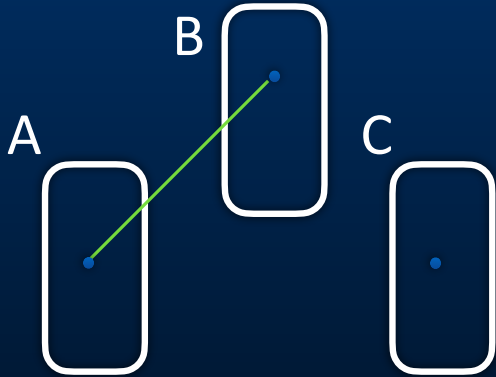
Want: Node gadgets such that: $OPT(\text{blue } a, \text{red } b) > \tau$

iff
 $(a, b) \in E$



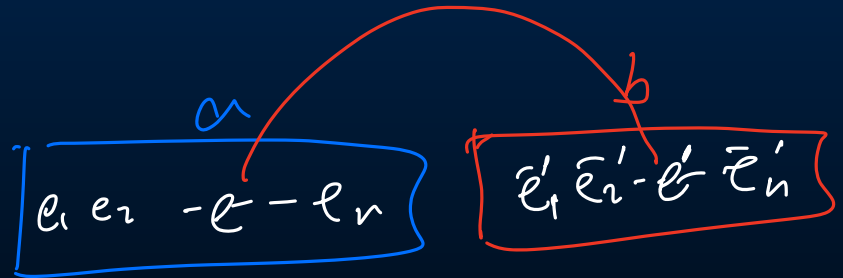
Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that: $OPT(\text{blue } a, \text{red } b) > \tau$



iff
 $(a, b) \in E$

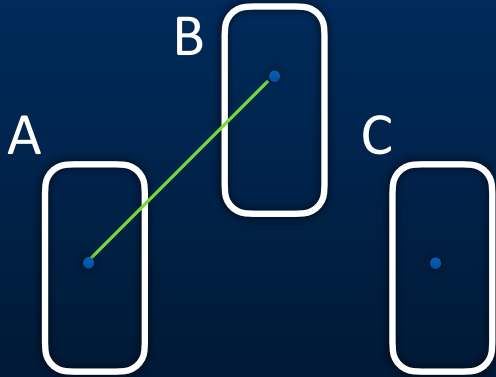
How?



1 if \exists common edge
 0 otherwise.

Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that:



$$OPT(\text{blue } a \text{ red } b) > \tau$$

iff

$$(a, b) \in E$$

How?

a

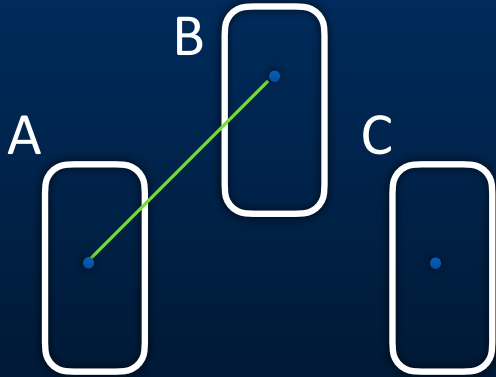
Just write a

$N_A(b) : a_3, a_7, a_{20}, \dots$

List all $a \in N_A(b)$

Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that:



$$OPT(\text{blue } a \text{ red } b) > \tau$$

iff

$$(a, b) \in E$$

How?

a

$$N_A(b) : a_3, a_7, a_{20}, \dots$$

Just write a

List all $a \in N_A(b)$

$$OPT = 1 \text{ iff } (a, b) \in E$$

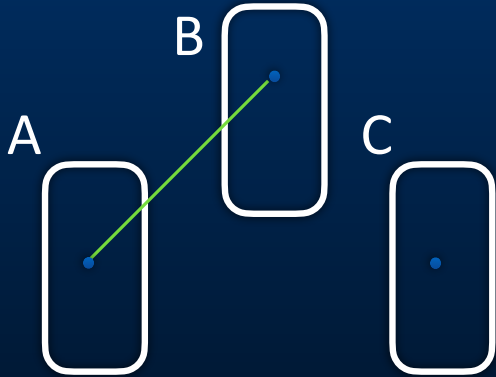
Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that:

$$OPT(\text{blue } a \text{ red } b) > \tau$$

iff

$$(a, b) \in E$$



How?

a

Just write a

$N_A(b) : a_3, a_7, a_{20}, \dots$

List all $a \in N_A(b)$

$$OPT = 1 \text{ iff } (a, b) \in E$$

Issue: Gadget size is $\Theta(n)$... Total length is $\Theta(n^2)$.

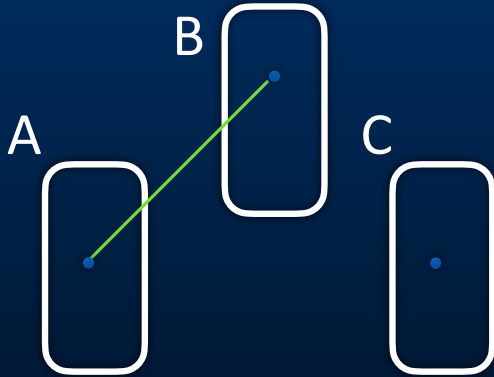
Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that:

$$OPT(\text{blue } a \text{ red } b) > \tau$$

iff

$$(a, b) \in E$$



How?

$$a$$

$$N_A(b) : a_3, a_7, a_{20}, \dots$$

Just write a

List all $a \in N_A(b)$

$$OPT = 1 \text{ iff } (a, b) \in E$$

$$N = n^2$$

$N^{w/2-\epsilon} \Rightarrow n^{w-\epsilon}$ for Δ .

Issue: Gadget size is $\Theta(n)$... Total length is $\Theta(n^2)$.

Thm: $O(n^{\omega/2-\epsilon})$ for RNA Folding, breaks the Triangle bound.

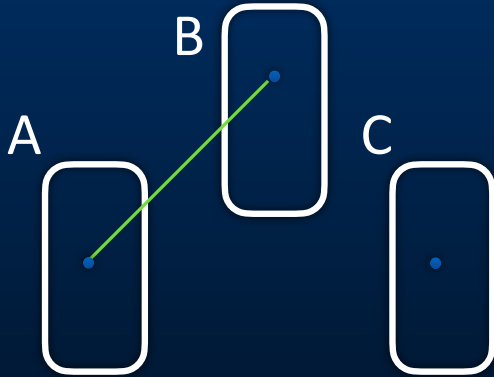
Higher Lower Bounds? $\Omega(n^\omega)$?

Want: Node gadgets such that:

$$OPT(\text{blue } a \text{ red } b) > \tau$$

iff

$$(a, b) \in E$$



a

Just write *a*

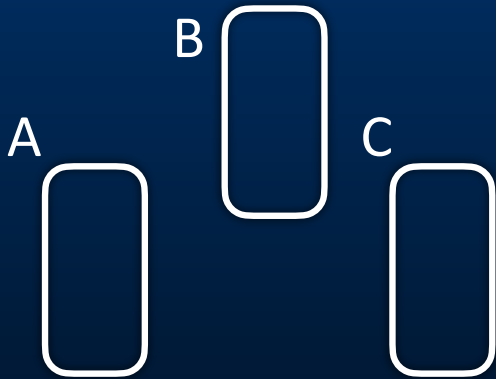
$N_A(b) : a_3, a_7, a_{20}, \dots$

List all $a \in N_A(b)$

Want: Shorter node gadgets... $\tilde{O}(1)$ bits.

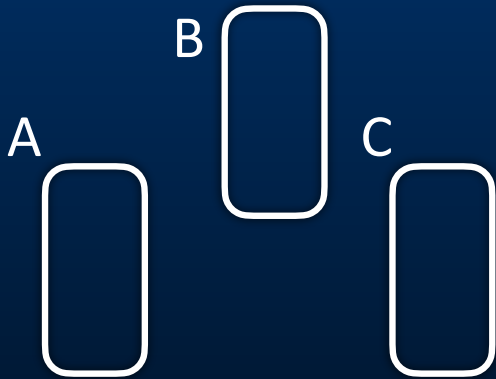
Issue: Impossible... need $\Omega(n^2)$ bits to represent a graph!

Is Triangle hard on “compressible” graphs?



Can the edges be encoded
with $o(n^2)$ bits?

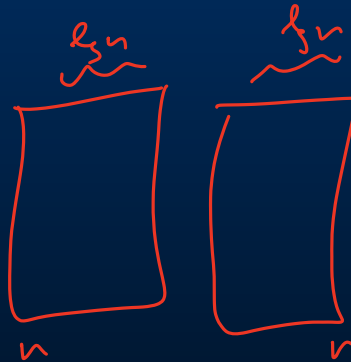
Is Triangle hard on “compressible” graphs?



Can the edges be encoded
with $o(n^2)$ bits?

k -SAT
 n^k bits \rightarrow OV
 $N = 2^{n/2}$
 $(0.5N)^k$

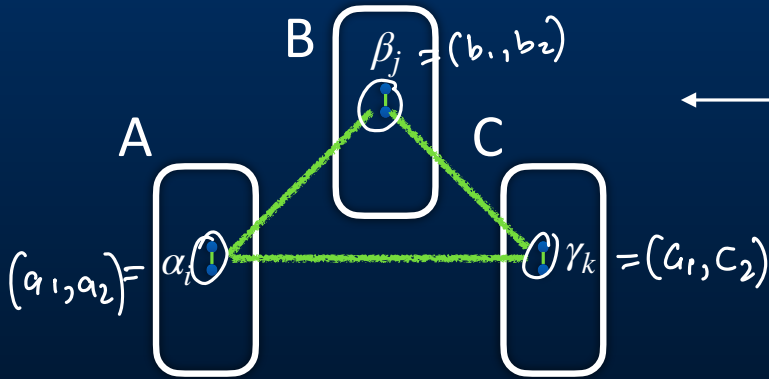
Is OV hard on “compressible” inputs?



$\tilde{O}(n)$ bits in general!
SETH $\Rightarrow \Omega(n^2)$ even
when the instance
can be represented
in $\tilde{O}(1)$ bits.

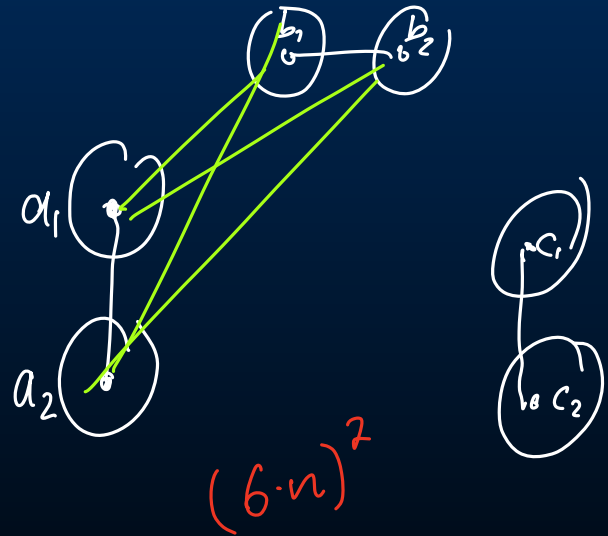
Is Triangle hard on “compressible” graphs?

3-Clique



6-Clique

$O(n^{2\omega})$ time



$N = O(n^2)$ nodes,

$O(N^2) = O(n^4)$ edges,

But only $O(N) = O(n^2)$ bits!

$\Omega(n^\omega)$

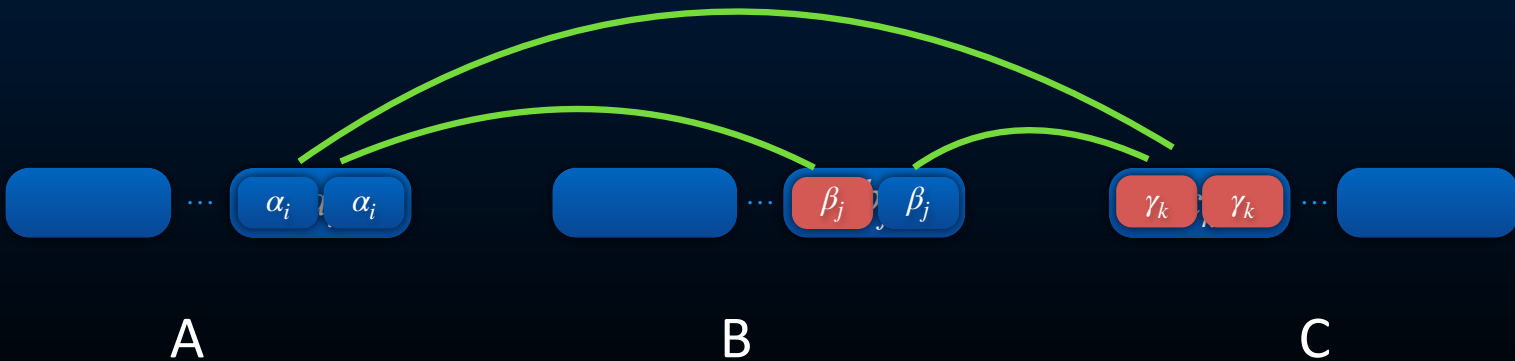
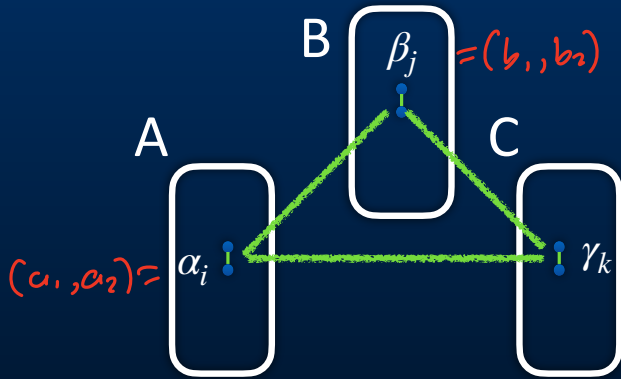
k-Clique Conjecture: “yes”

Higher Lower Bounds? $\Omega(n^\omega)$?

Want: “Node” gadgets such that: $OPT(\alpha = a_1, a_2 \quad \beta = b_1, b_2) > \tau$

iff

$$(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2) \in E$$

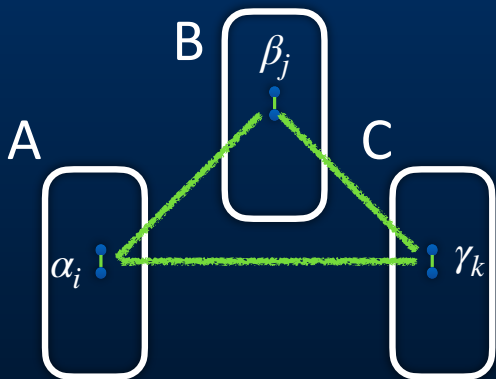


Higher Lower Bounds? $\Omega(n^\omega)$?

Want: “Node” gadgets such that: $OPT(\alpha = a_1, a_2 \quad \beta = b_1, b_2) > \tau$

iff

$$(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2) \in E$$



$$a_1 \quad a_2 \quad a_2 \quad a_1$$

$$N(b_1) \quad N(b_2) \quad N(b_1) \quad N(b_2)$$

$$\# a_1 \# a_2 \# a_1 \# a_2 \#$$

$$\# N(b_1) \# N(b_1) \# N(b_2) \# N(b_2) \#$$

$A_2 \quad A_1$

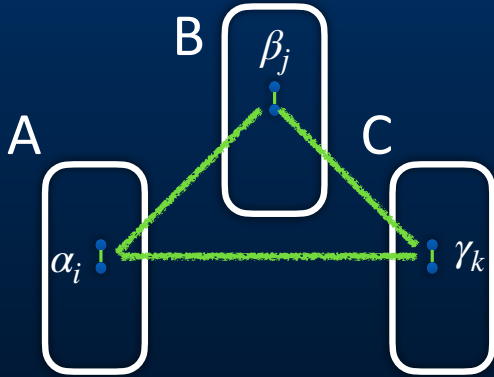
$OPT = 4$ iff (α, β) is a 2x2-biclique.

k-Clique Lower Bounds $\Omega(n^\omega)$

Want: “Node” gadgets such that: $OPT(\alpha = a_1, a_2 \quad \beta = b_1, b_2) > \tau$

iff

$$(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2) \in E$$



$N = n^3 \leftarrow \Omega(n^{2\omega})$ *6-Clique*

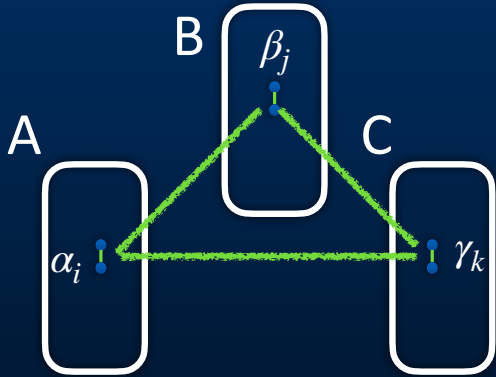
$\Omega(N^{\omega \cdot \frac{2}{3} - \epsilon})$

Total length is $O(n^2) \cdot O(n) = \underline{O(n^3)}$.

Thm: $O(n^{2\omega/3 - \epsilon})$ for RNA Folding, breaks the 6-Clique bound.

k-Clique Lower Bounds $\Omega(n^\omega)$

Want: “Node” gadgets such that: $OPT(\alpha = a_1, a_2 \quad \beta = b_1, b_2) > \tau$



iff

$$(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2) \in E$$

$N = n^k \cdot n = n^{k+1}$ 3k-clique
 $= n^{k+1} \leftarrow \Omega(n^{\omega k})$

Total length is $O(n^2) \cdot O(n) = O(n^3)$.

$$\Omega\left(\frac{N^k}{k+1} \cdot \omega - \varepsilon\right)$$

Thm: $O(n^{2\omega/3 - \varepsilon})$ for RNA Folding, breaks the 6-Clique bound.

Start from 3k-Clique: total length is $O(n^{k+1})$.

Thm: $O(n^{\frac{k}{k+1}\omega - \varepsilon})$ for RNA Folding, breaks the 3k-Clique bound.

k-Clique Conjecture(s)

Weighted k-Clique Conjecture:

No $O(n^{(1-\varepsilon)\cdot k})$ time alg for Min-Weight k-Clique.

Combinatorial k-Clique Conjecture:

No $O(n^{(1-\varepsilon)\cdot k})$ time alg for k-Clique, combinatorially.

Unweighted k-Clique Conjecture:

No $O(n^{(\omega/3-\varepsilon)\cdot k})$ time alg for k-Clique.

k-Clique

Longstanding Open Questions:

- Faster, $O(n^{(\omega/3-\varepsilon)\cdot k})$ for k-Clique?
- $O(n^{(1-\varepsilon)\cdot k})$ without matrix mult.?
- $O(n^{(1-\varepsilon)\cdot k})$ for Min-Weight-k-Clique?

Some Known Hardness for k-Clique:

The canonical **W[1]-complete** problem.

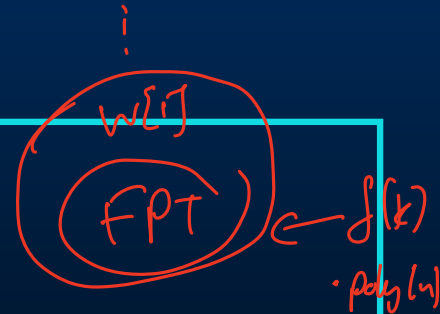
$n^{o(k)}$ refutes ETH. [Chen et al. 05]

$O(n^{k/2-\varepsilon})$ implies faster **k-SUM**. [A.-Lewi-Williams 14]

Faster alg implies faster **MAX-CUT**. [Williams 04]

etc...

MAX-2SAT 1.21^n



\sqrt{k}

RNA Folding

Upper bounds:

k-Clique lower bounds:

Dynamic Programming

$$O(n^3)$$

$$\Omega(n^{3-\varepsilon}) \text{ (Combinatorial)}$$

$$\Omega(n^{\omega-\varepsilon})$$

RNA Folding

Upper bounds:

k-Clique lower bounds:

Dynamic Programming

$$O(n^3)$$

$$\Omega(n^{3-\varepsilon}) \text{ (Combinatorial)}$$

$O(n^{2.99})$ via MM?

$$\Omega(n^{\omega-\varepsilon})$$

[ABV FOCS '15]

RNA Folding

Upper bounds:

k-Clique lower bounds:

Dynamic Programming

$$O(n^3)$$

$$\Omega(n^{3-\varepsilon}) \text{ (Combinatorial)}$$

$O(n^{2.99})$ via MM?

$$\Omega(n^{\omega-\varepsilon})$$

[ABV FOCS '15]

CFG Parsing

CYK Parser (60's)

(Combinatorial)

$$O(n^3)$$

$$\Omega(n^{3-\varepsilon}) \text{ (Combinatorial)}$$

Valiant's Parser ('75)

(via MM)

$$O(n^\omega)$$

$$\Omega(n^{\omega-\varepsilon})$$

RNA Folding

Upper bounds:

k-Clique lower bounds:

Dynamic Programming

$$O(n^3)$$

$$\Omega(n^{3-\varepsilon}) \text{ (Combinatorial)}$$

$O(n^{2.99})$ via MM?

$$O(n^{2+\omega/3})$$

[BGSV FOCS'16, CDX SODA'22]

$$\Omega(n^{\omega-\varepsilon})$$

[ABV FOCS '15]

CFG Parsing

CYK Parser (60's)

(Combinatorial)

$$O(n^3)$$

$$\Omega(n^{3-\varepsilon}) \text{ (Combinatorial)}$$

Valiant's Parser ('75)

(via MM)

$$O(n^\omega)$$

$$\Omega(n^{\omega-\varepsilon})$$

Tight lower bounds under the k-Clique conjectures

[A. - Backurs - Vassilevska W FOCS'15]

- CFG Parsing
- RNA Folding
- Dyck Language Edit Distance

Formal Languages.

Computational Biology.

[Backurs - Dikkala - Tzamos ICALP'16]

- Max Weight Rectangle

Comp Geometry.

[Backurs - Tzamos ICML'17]

- Improving Viterbi's Algorithm

Machine Learning.

[Bringmann - Gronlund - Larsen FOCS'17]

- Word Break

Pattern Matching.

[Bringmann - Gawrychowski - Mozes - Weimann SODA'18]

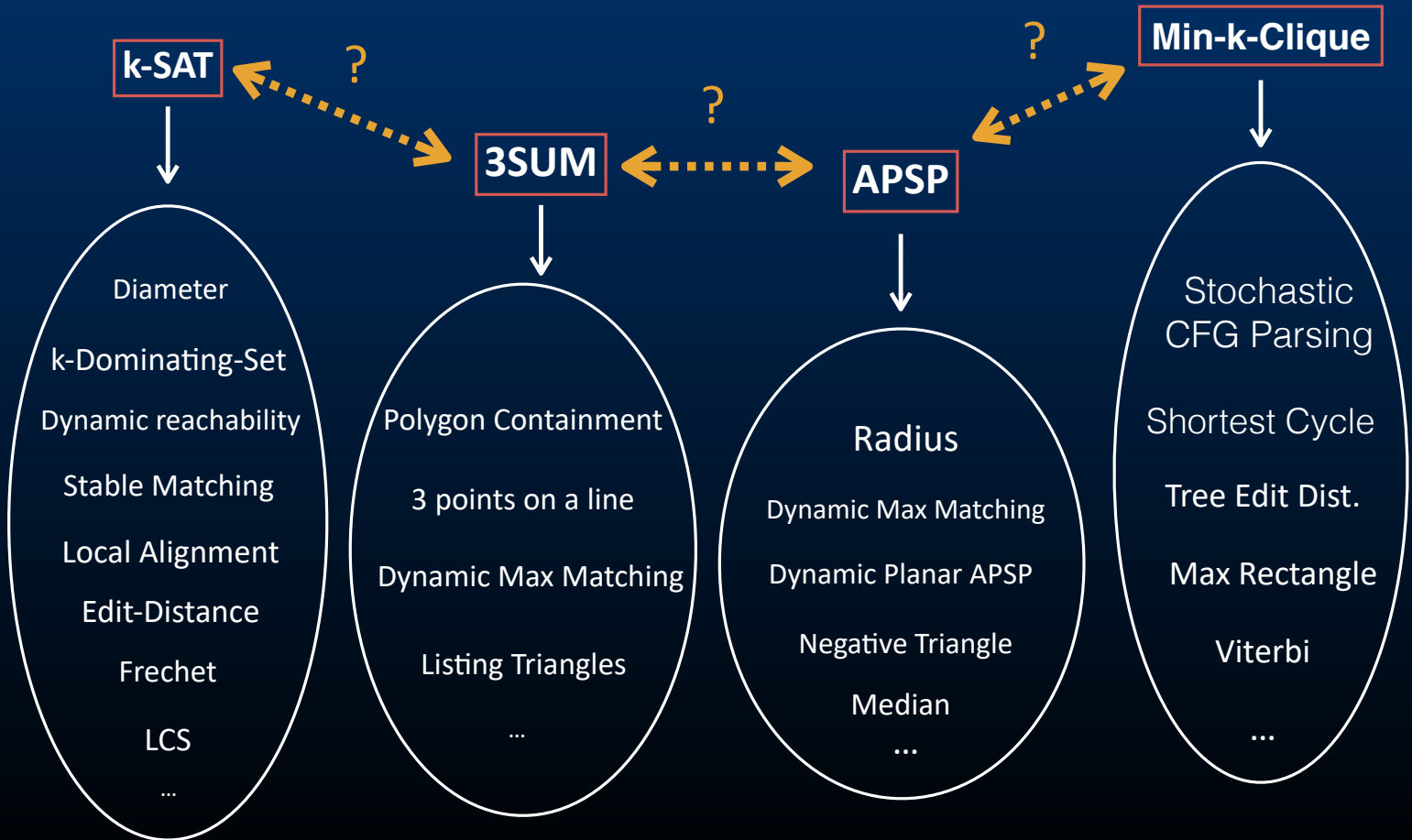
- Tree Edit Distance

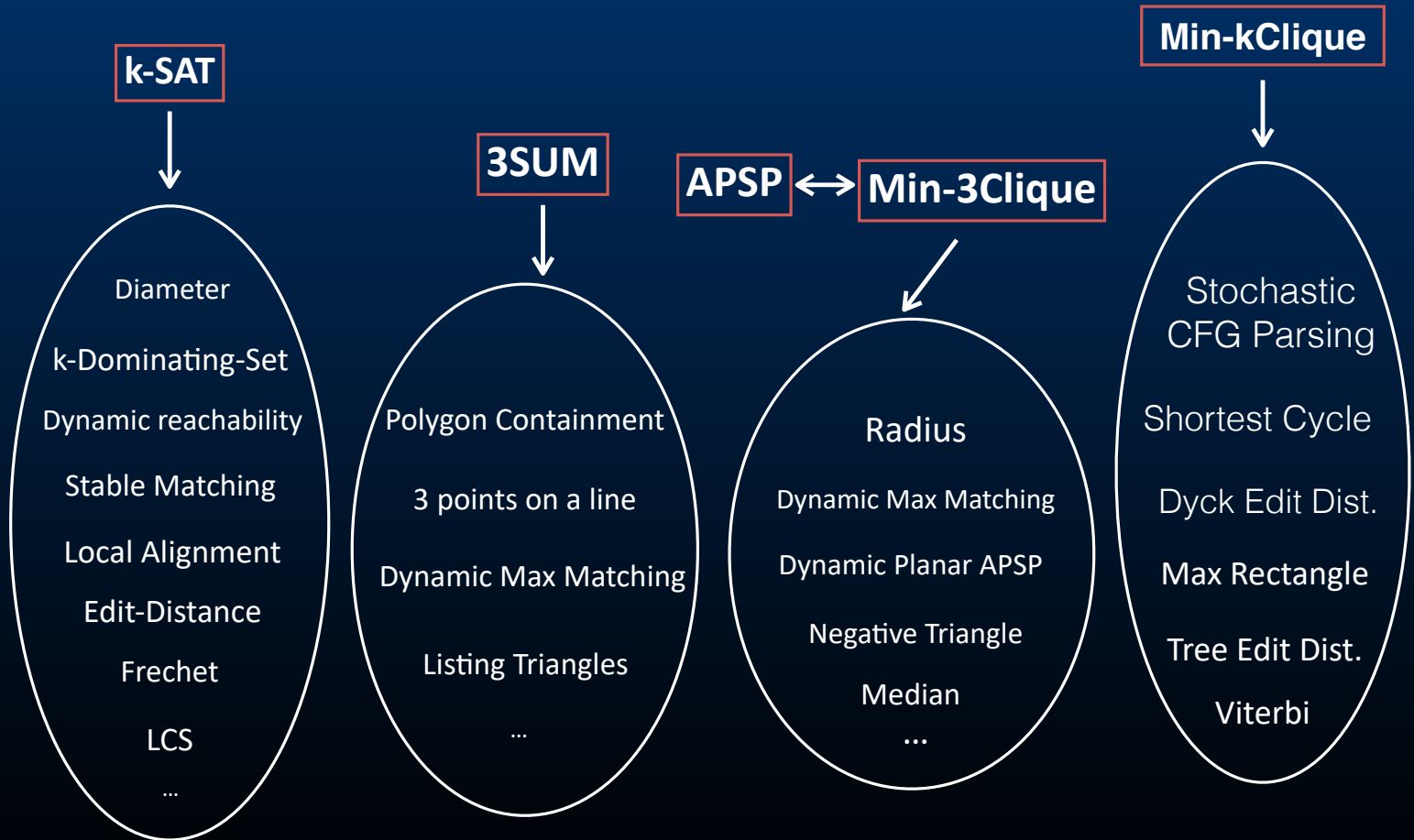
[Lincoln - Vassilevska W - Williams SODA'18]

- Shortest Cycle in Sparse Graphs

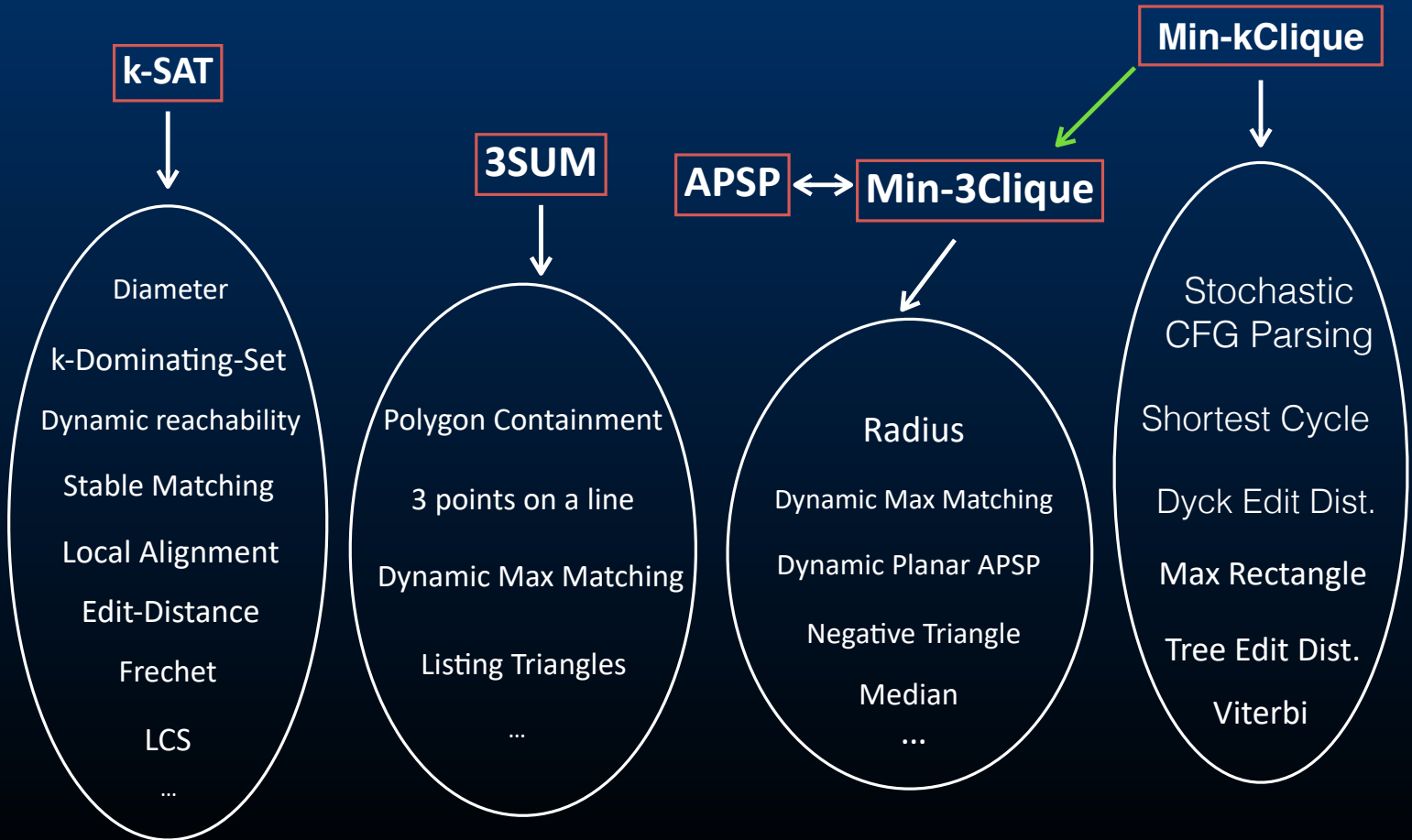
Distance computations.

How are these conjectures related?





An easy connection



Barrier: low complexity to high complexity

300-Clique in $n^{300-\epsilon}$ \Rightarrow 3-Clique in $n^{3-\epsilon}$
 $N = n^{\frac{1}{100}}$

k-SAT



- Diameter
- k-Dominating-Set
- Dynamic reachability
- Stable Matching
- Local Alignment
- Edit-Distance
- Frechet
- LCS
- ...

3SUM



- Polygon Containment
- 3 points on a line
- Dynamic Max Matching
- Listing Triangles
- ...

APSP

Min-3Clique



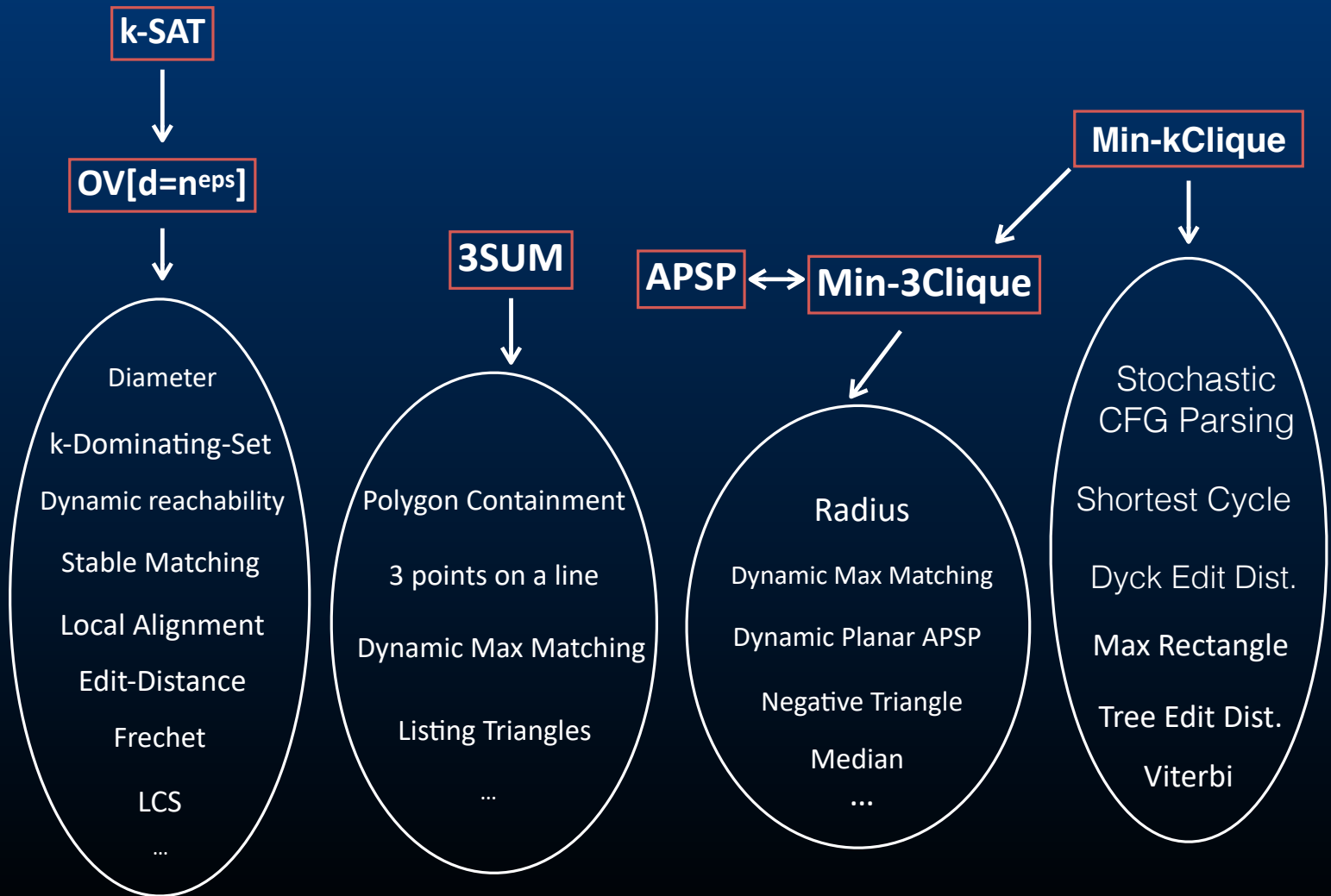
- Radius
- Dynamic Max Matching
- Dynamic Planar APSP
- Negative Triangle
- Median
- ...

Min-kClique

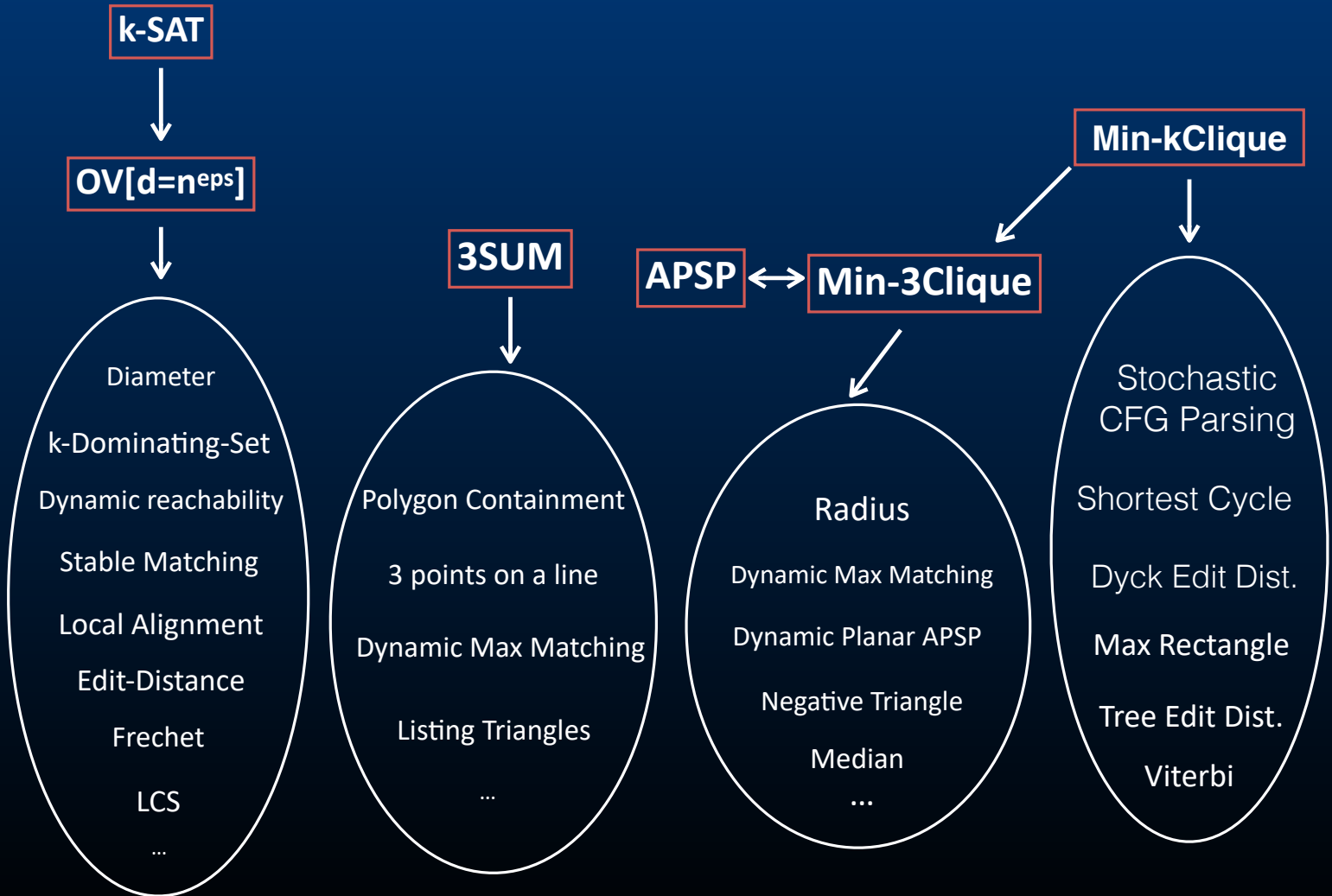


- Stochastic CFG Parsing
- Shortest Cycle
- Dyck Edit Dist.
- Max Rectangle
- Tree Edit Dist.
- Viterbi

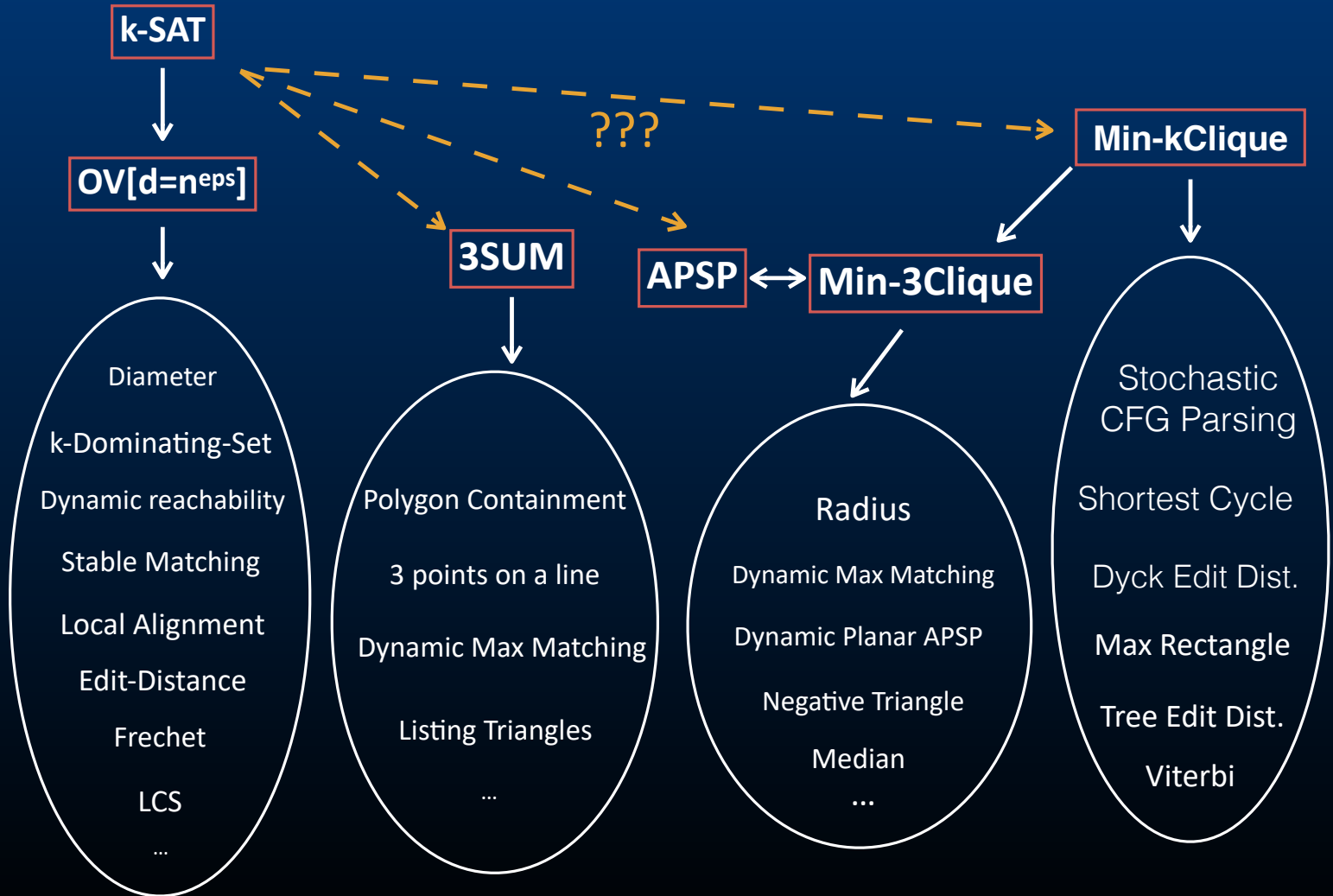




More connections?

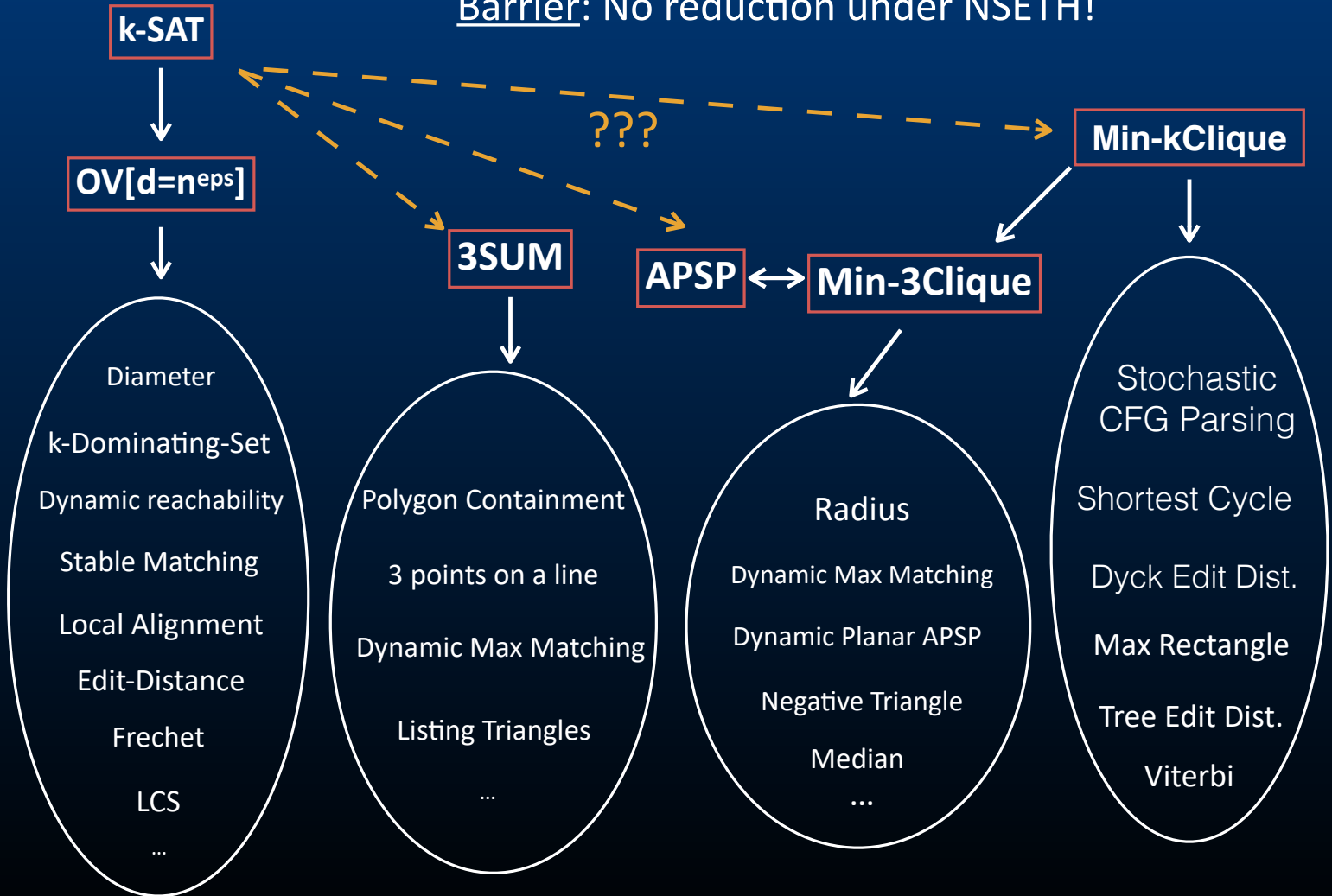


More connections?

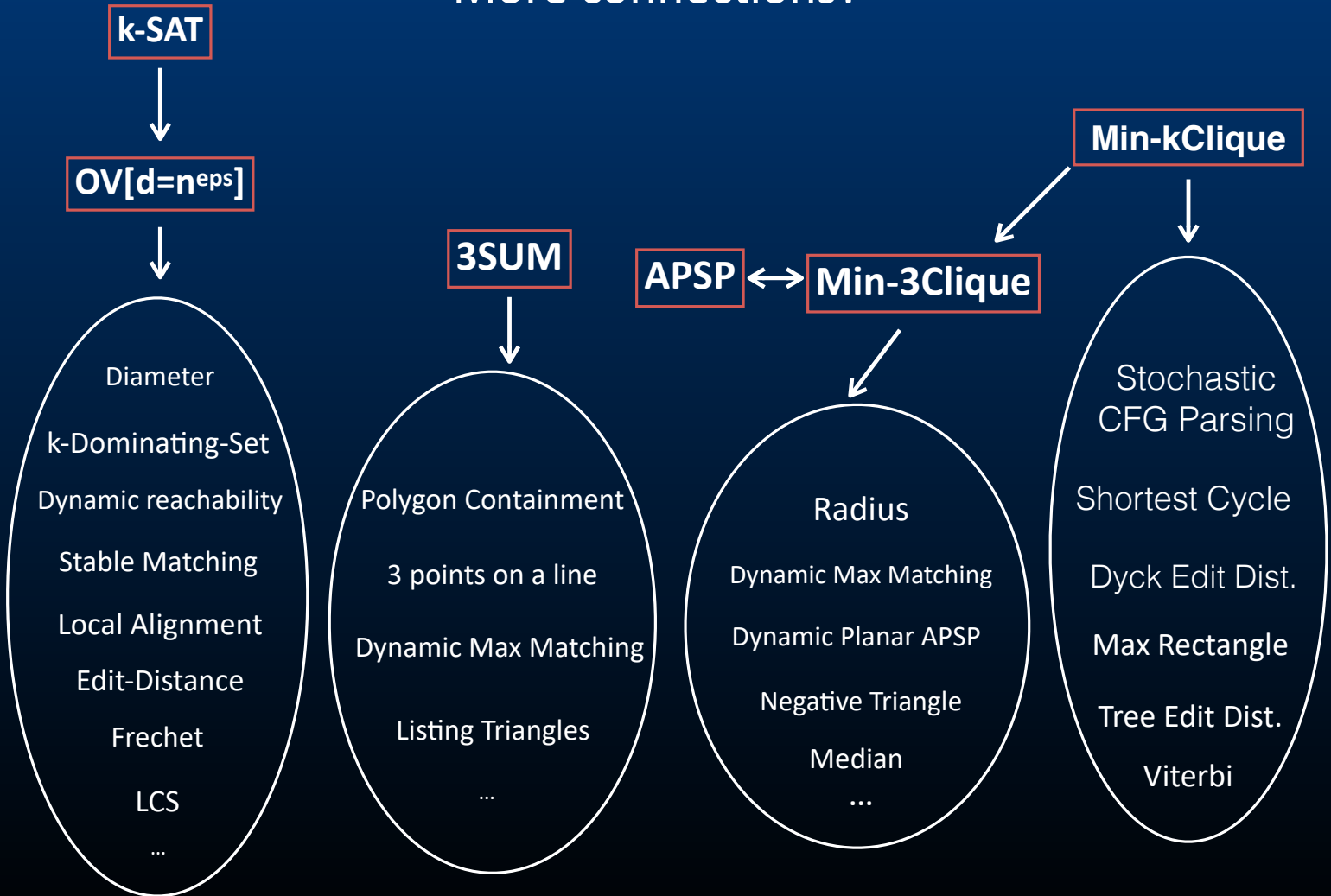


[Carmosino - Gao - Impagliazzo - Mikhailin - Paturi - Schneider '16]

Barrier: No reduction under NSETH!



More connections?



[A.-Bringmann-Dell-Nederlof '18]
Clique implies OV.

k-SAT



OV[d=neps]



- Diameter
- k-Dominating-Set
- Dynamic reachability
- Stable Matching
- Local Alignment
- Edit-Distance
- Frechet
- LCS
- ...

"weight removal" and other tricks



3SUM



- Polygon Containment
- 3 points on a line
- Dynamic Max Matching
- Listing Triangles
- ...

APSP

Min-3Clique



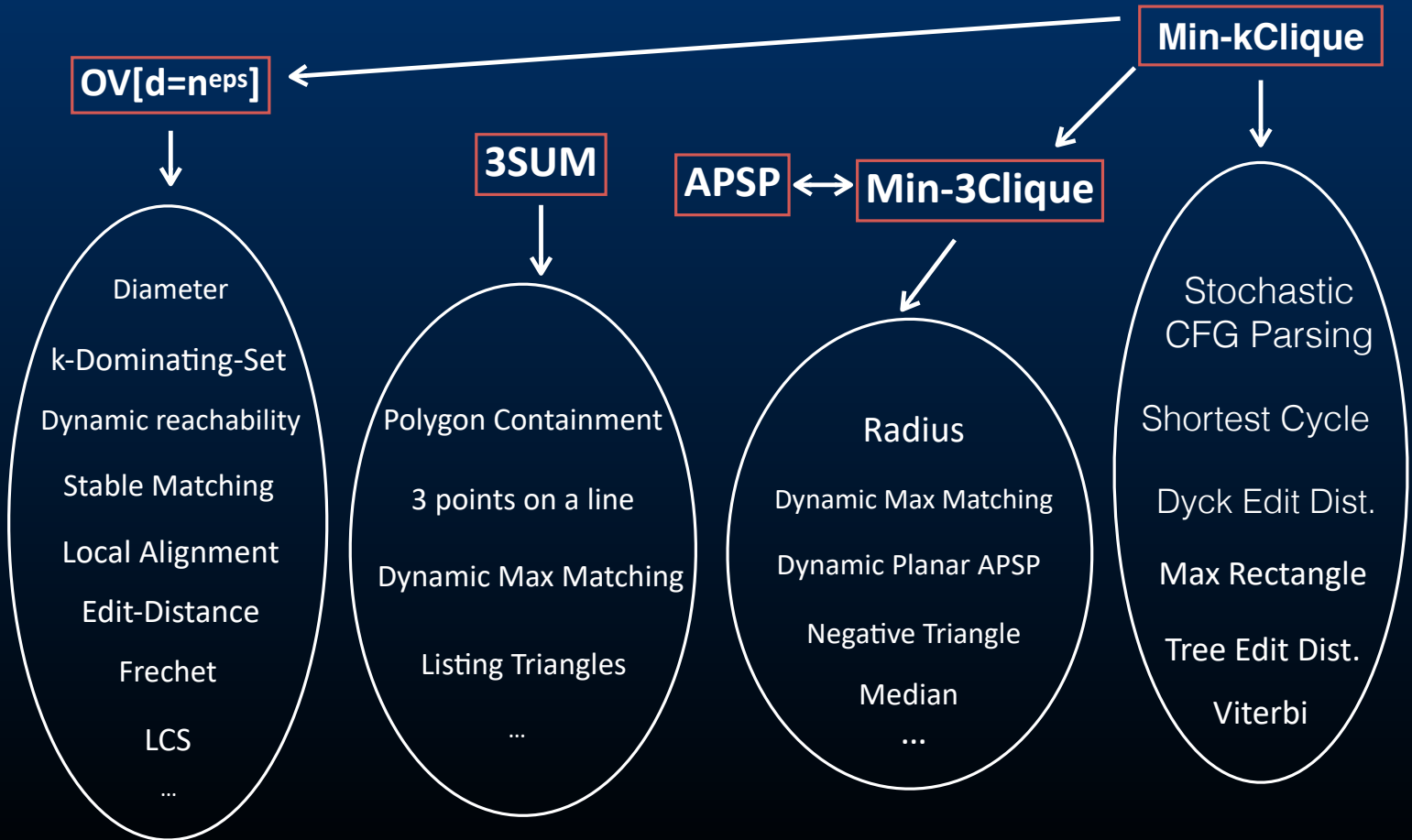
- Radius
- Dynamic Max Matching
- Dynamic Planar APSP
- Negative Triangle
- Median
- ...

Min-kClique



- Stochastic CFG Parsing
- Shortest Cycle
- Dyck Edit Dist.
- Max Rectangle
- Tree Edit Dist.
- Viterbi

The source of all lower bounds in P?



The source of all lower bounds in P?

