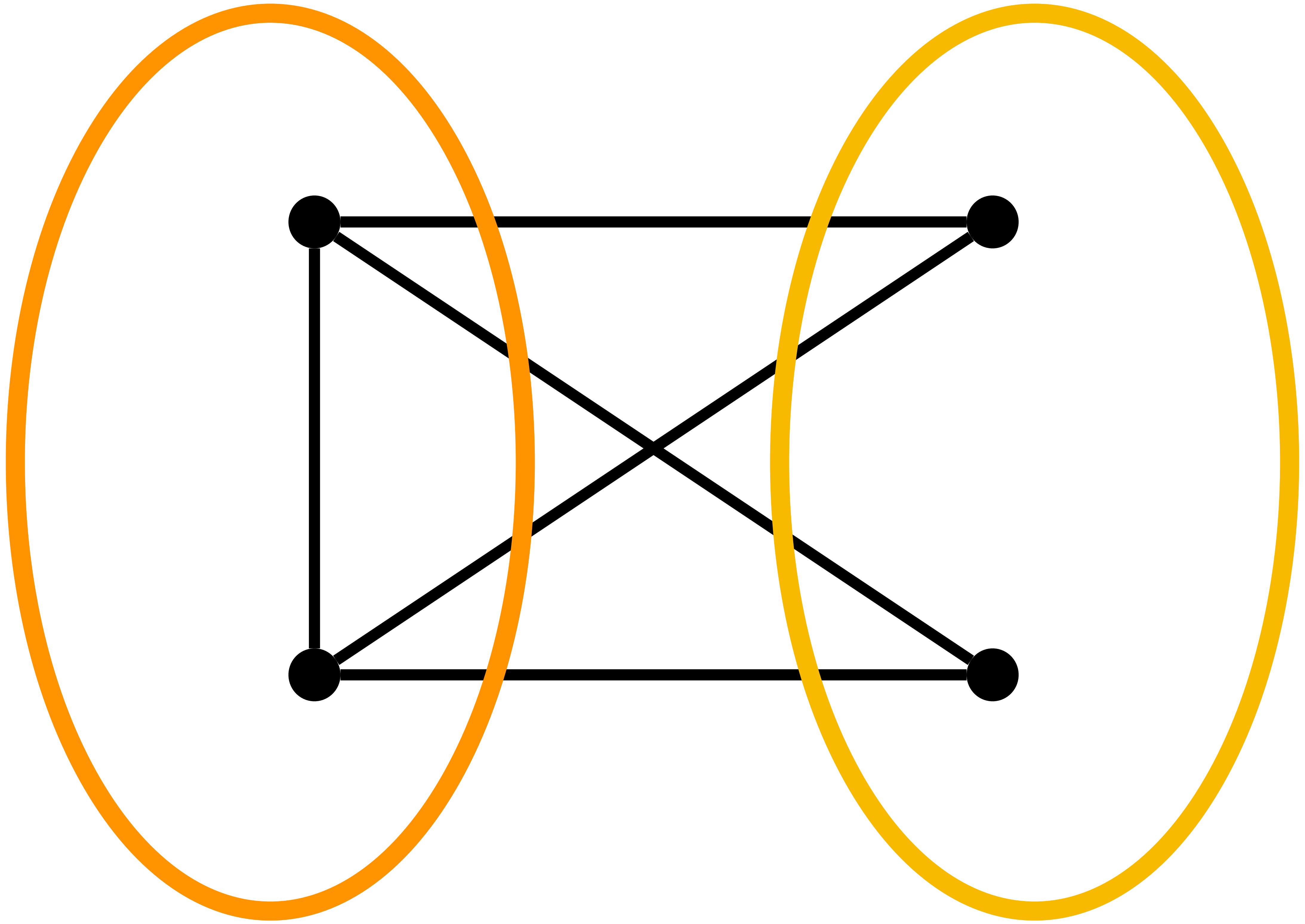
Sketching Algorithms for Maximum Directed Cut and other CSPs

Santhoshini Velusamy
Toyota Technological Institute at Chicago

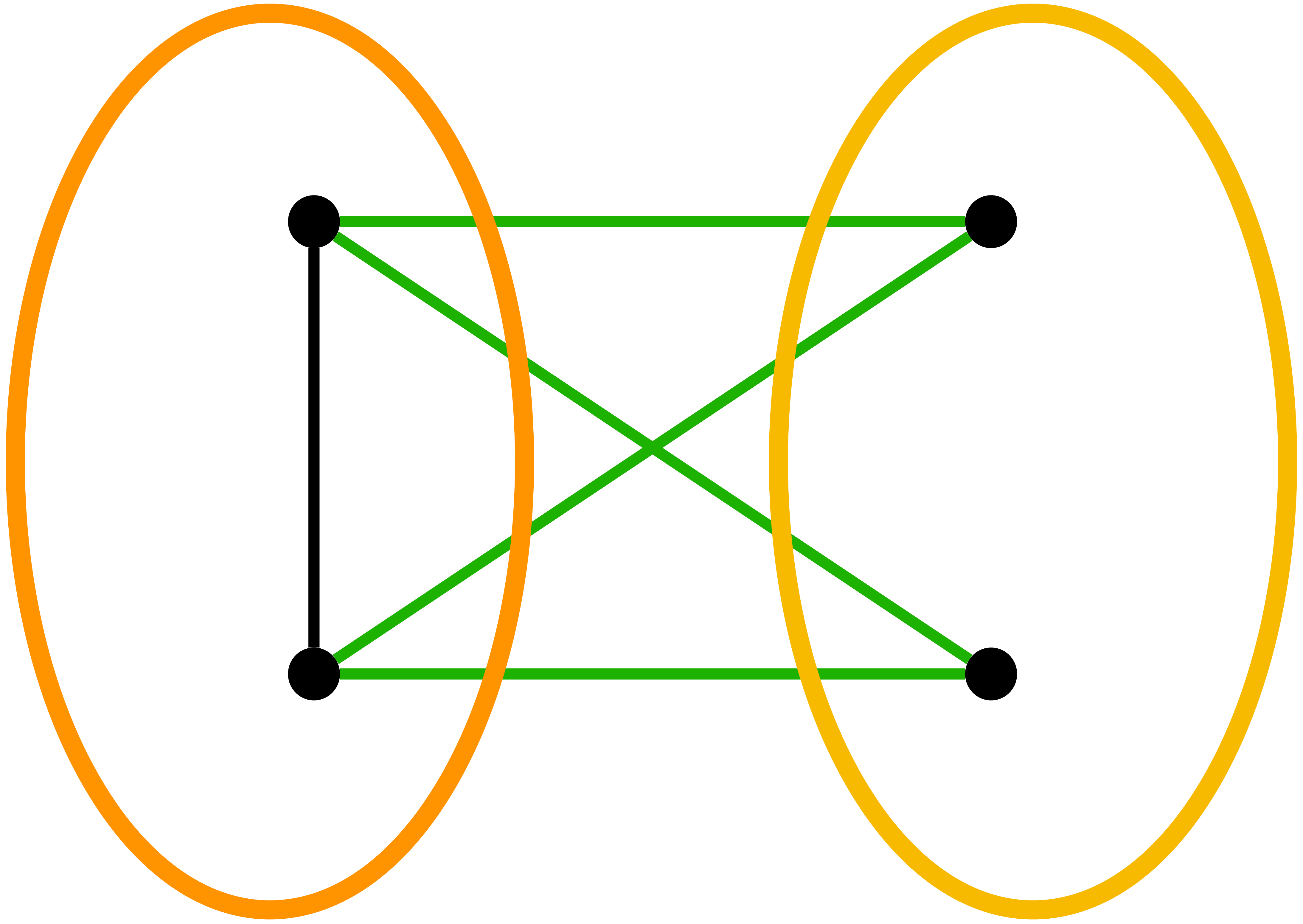
Talk Outline

- Problem definition and folklore algorithms
- **Sketching algorithms for Max-DICUT and lower bounds**
- Generalizations for arbitrary CSPs (brief)
- Open questions

Max-CUT



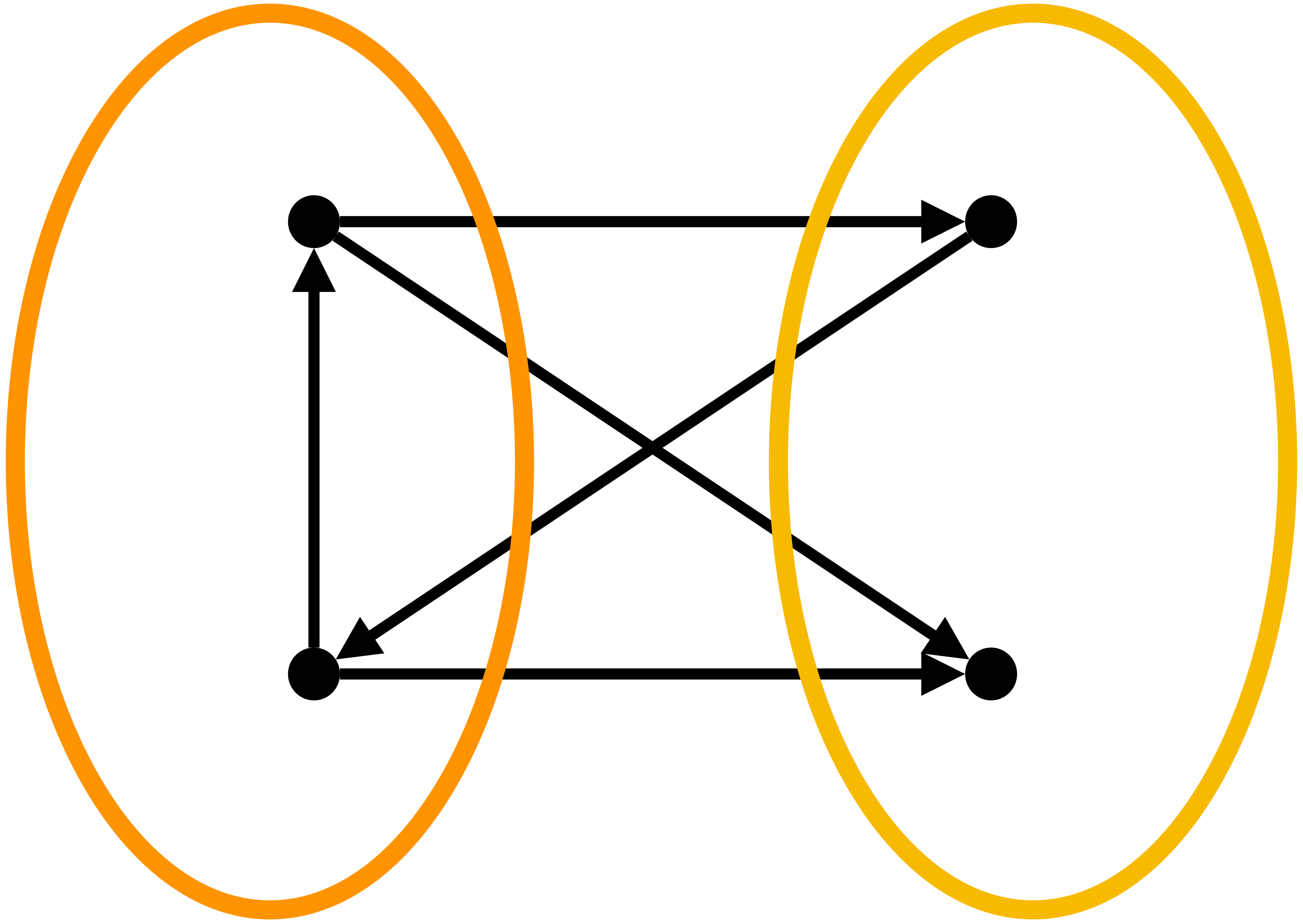
Max-CUT



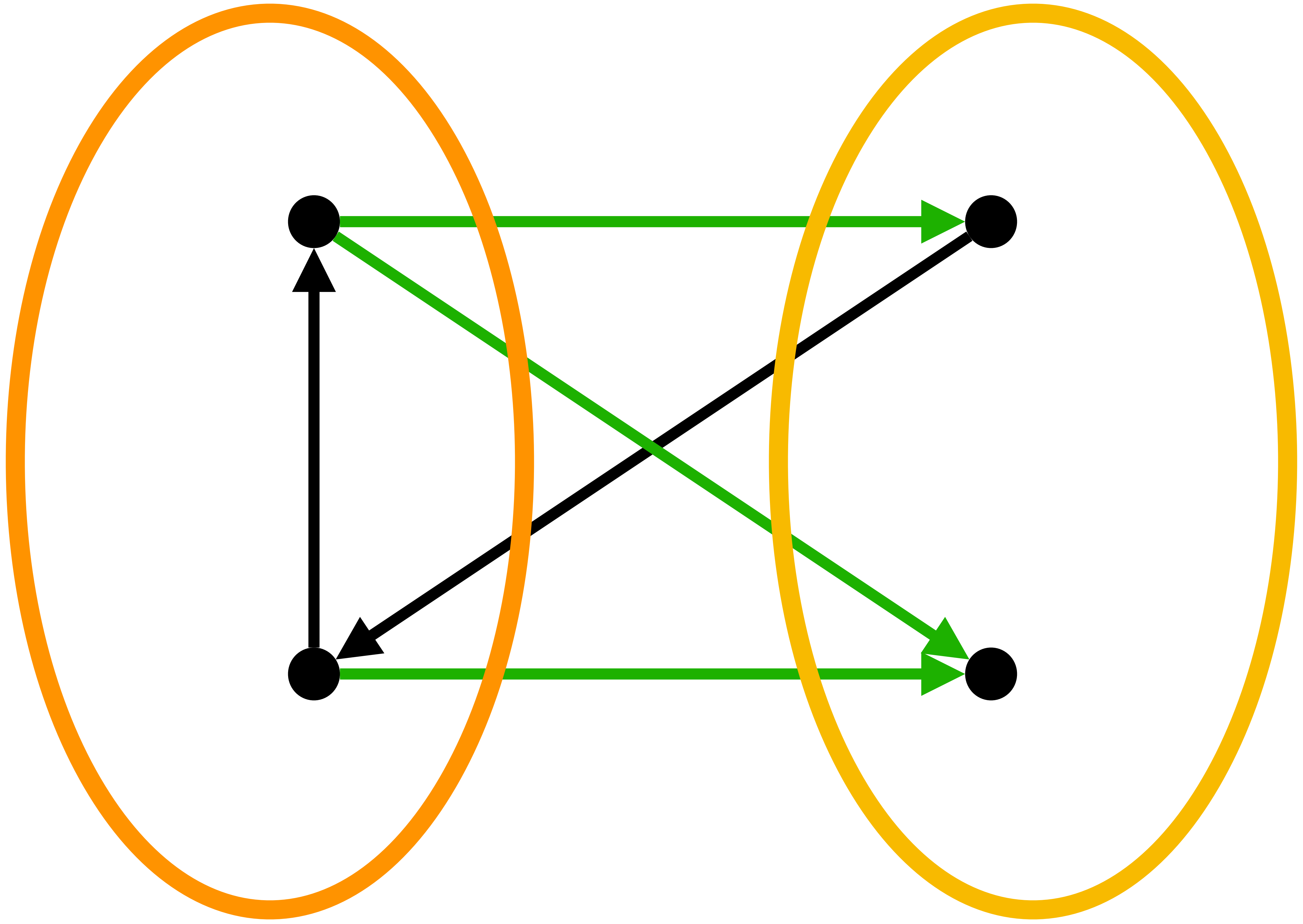
CUT size = 4

Maximize CUT size

Max-DICUT



Max-DICUT

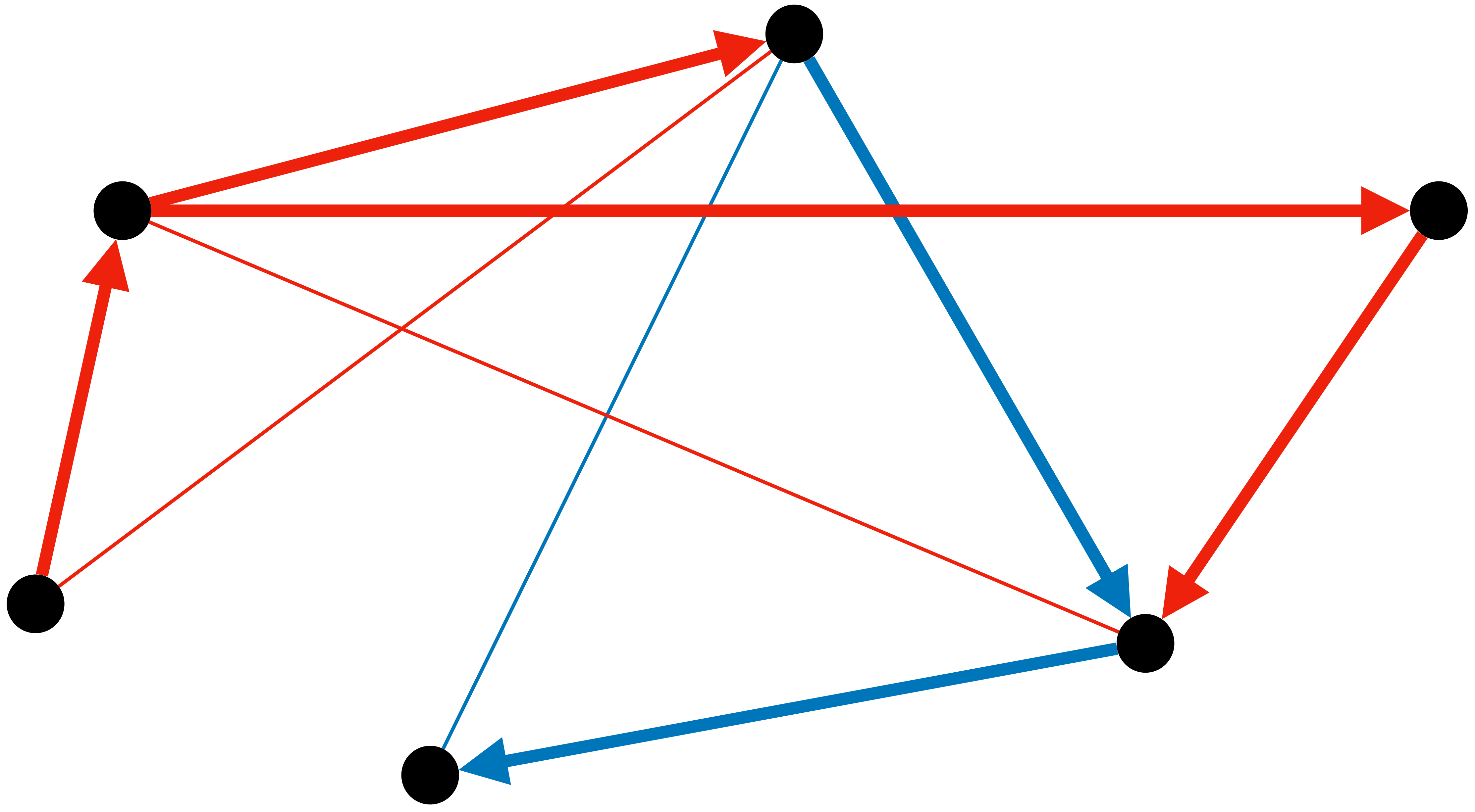


DICUT size = 3

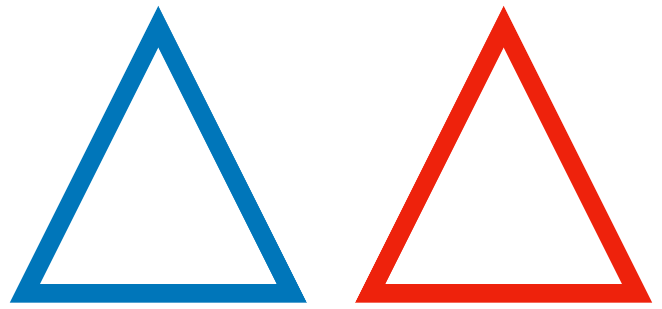
Maximize DICUT size

Constraint Satisfaction Problems (CSPs)

n **variables**

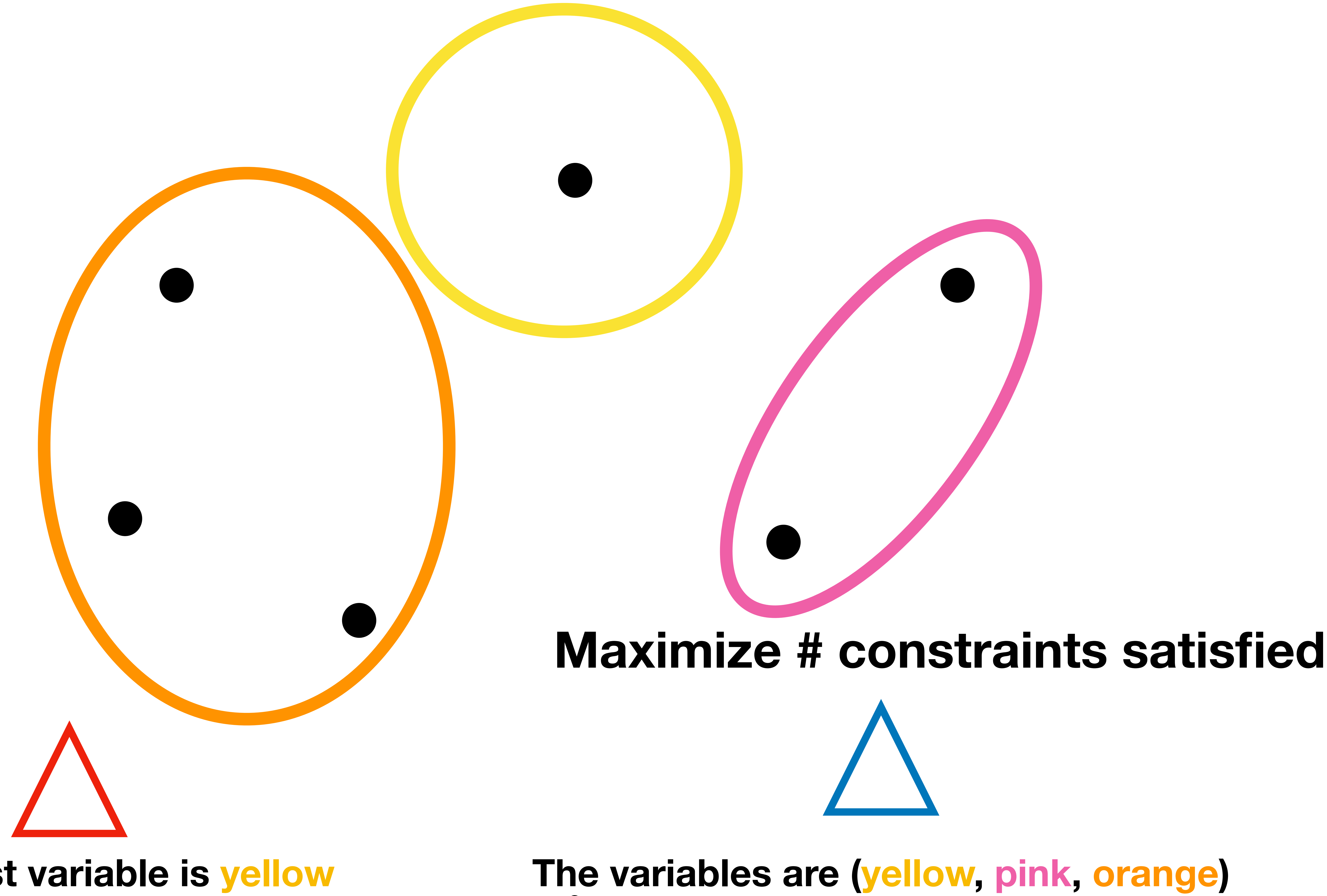


$k = 3$

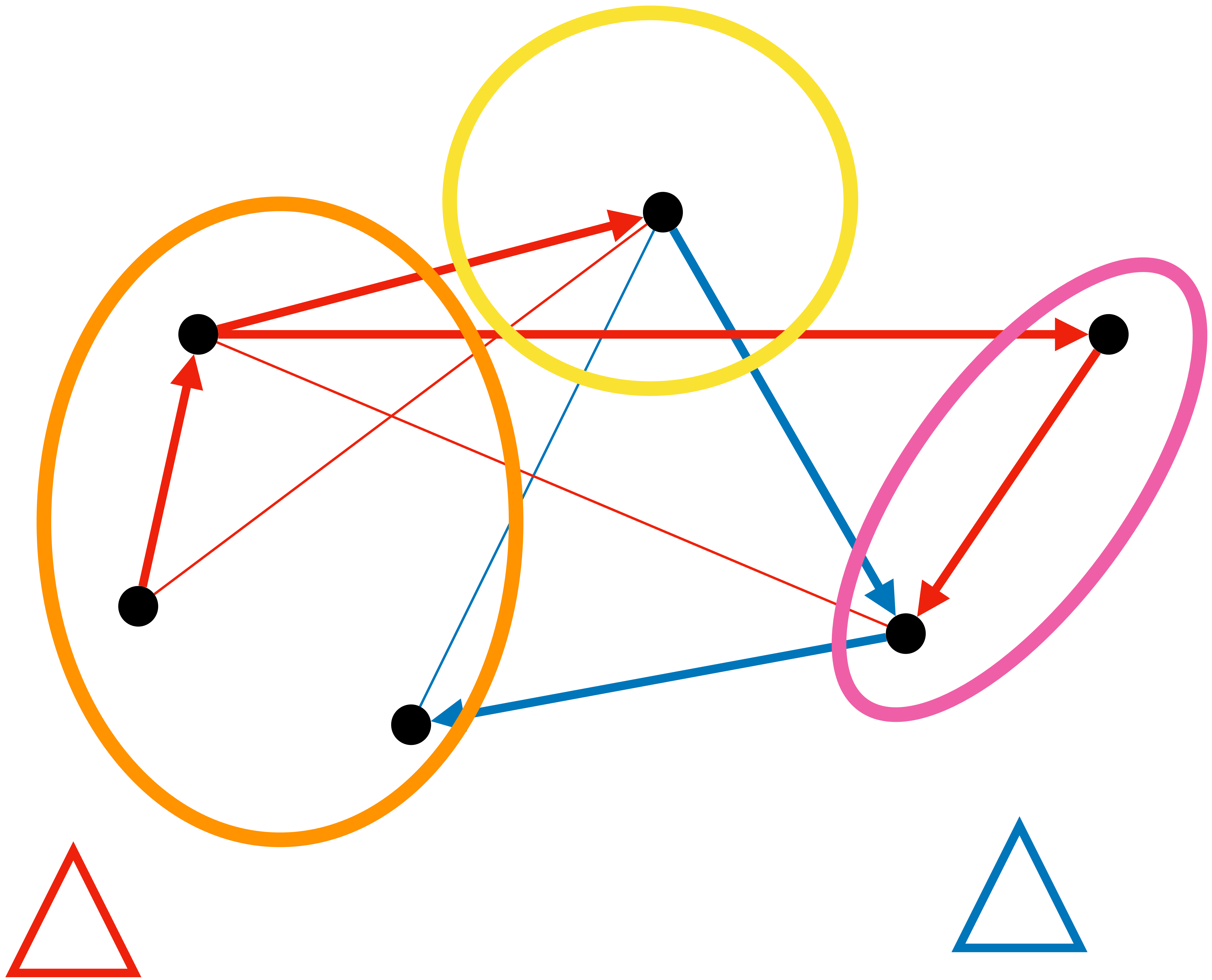


Constant number of constraint types of arity k

Constraint Satisfaction Problems (CSPs)



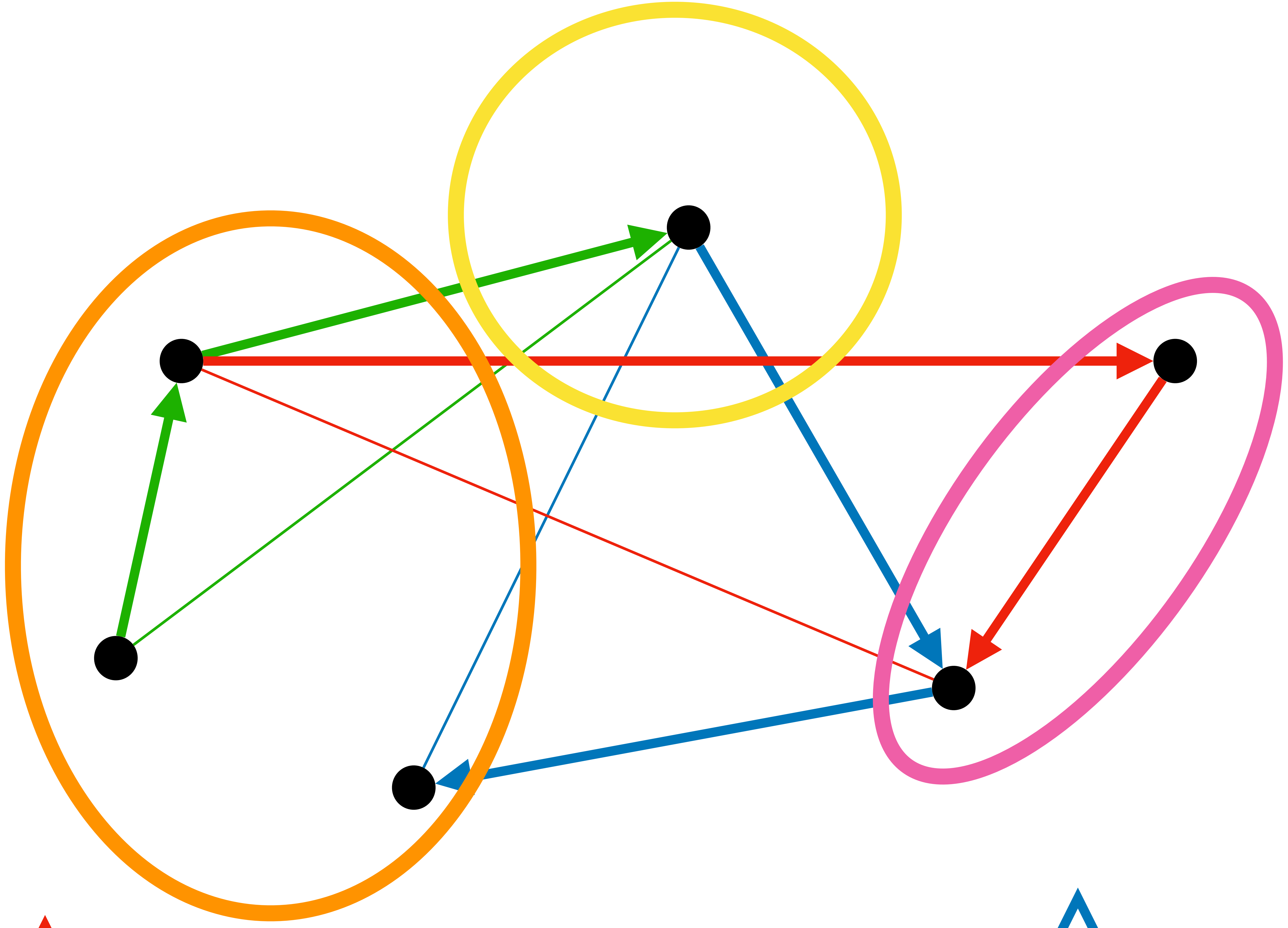
Constraint Satisfaction Problems (CSPs)

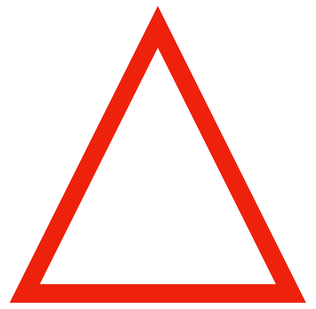


Last variable is yellow

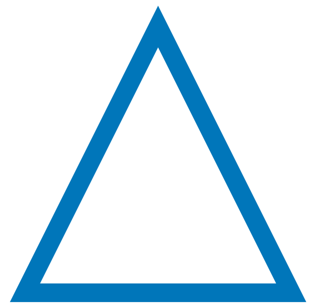
The variables are (yellow, pink, orange)

Constraint Satisfaction Problems (CSPs)



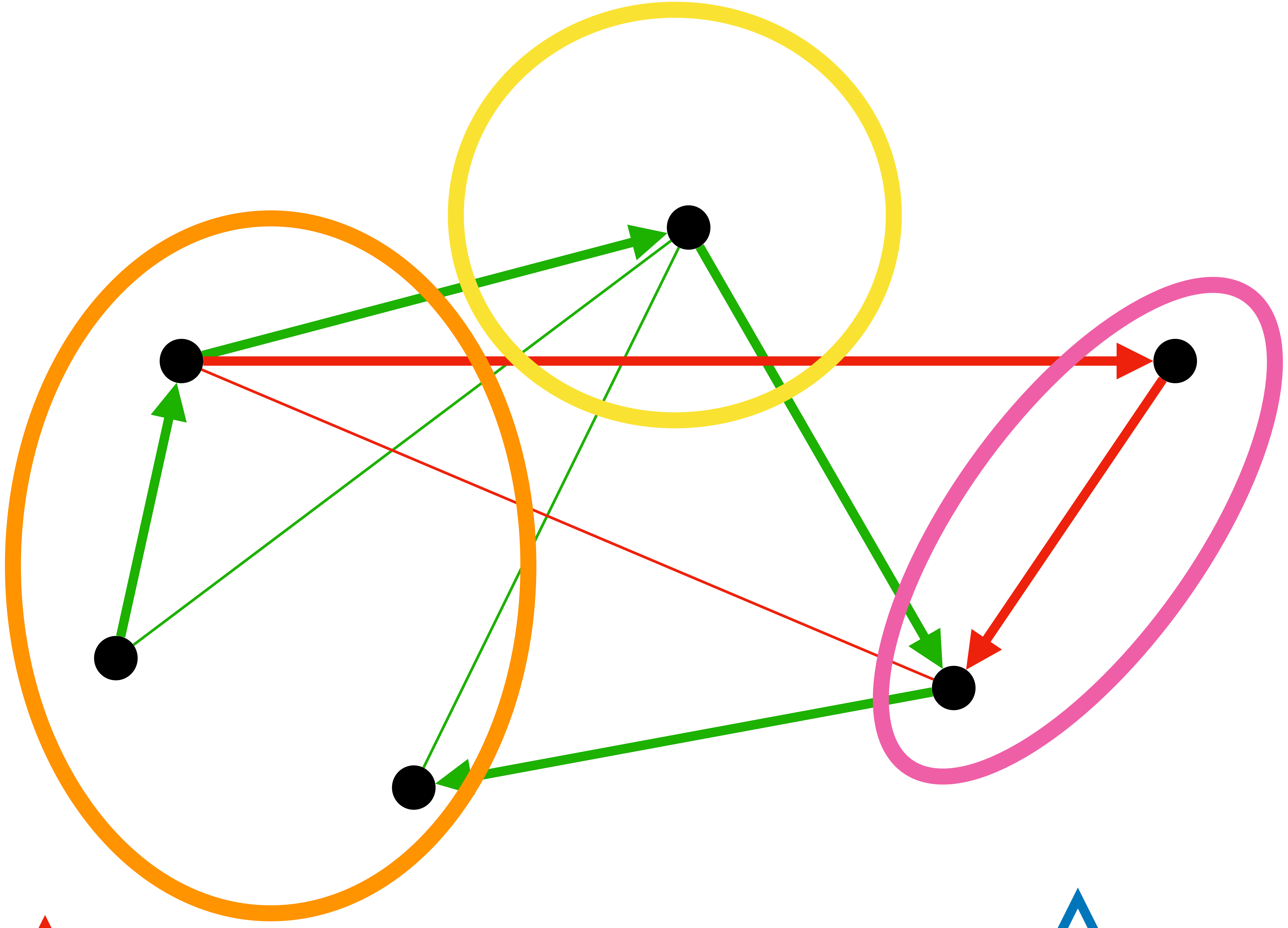


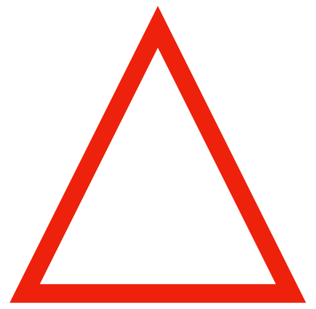
Last variable is yellow

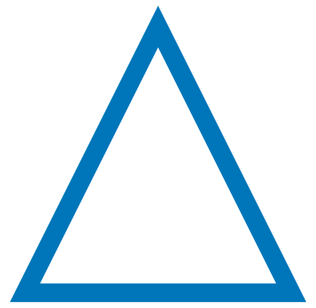


The variables are (yellow, pink, orange)

Constraint Satisfaction Problems (CSPs)







Last variable is yellow

The variables are (yellow, pink, orange)

Why study CSPs?

- General structure to study infinitely many problems simultaneously
- Allows for finite classifications through dichotomy theorems [Schaefer '78, Raghavendra '08, Khot-Tulsiani-Worah '14, Bulatov '17, Zhuk '20, Chou-Golovnev-Sudan-V ’21, Ghoshal-Lee '22, Kol-Parmanov-Saxena-Yu ’23]
- Discovery of new techniques that are broadly applicable. For example, SDP rounding, Sum of squares, Unique Games,..
- Helps understand the power and limitations of algorithms as a whole

Why study Max-DICUT?

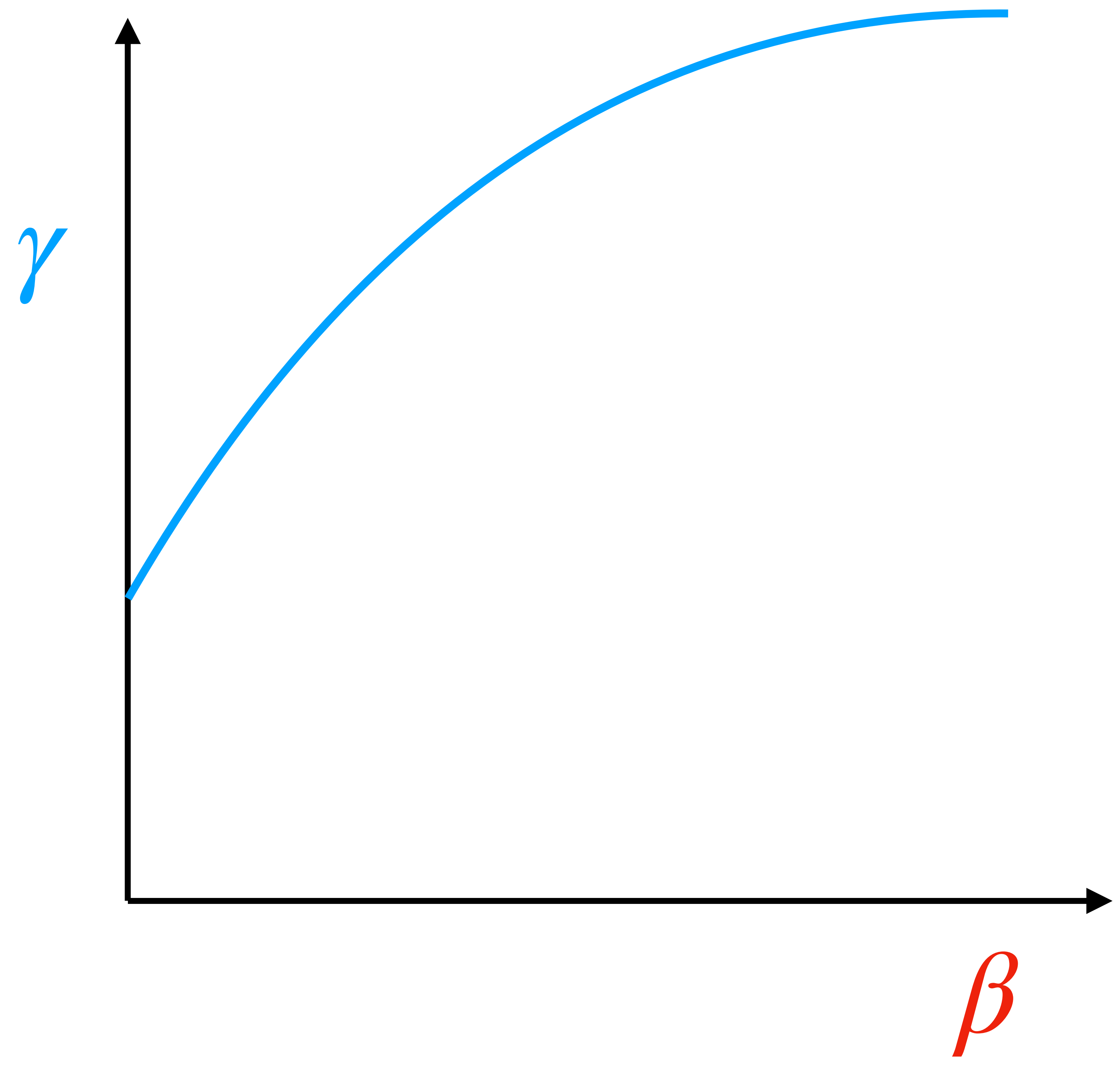
- Simplest problem with non-trivial approximability in the streaming setting
- [Chou-Golovnev-Sudan-V '21] Generalization of streaming algorithms and lower bounds of Max-DICUT to arbitrary CSPs

Approximation algorithms

- Exact computation of optimum maybe NP-Hard!
- $0 < \alpha < 1$, α approximation algorithm:
  - outputs T such that: $\alpha \cdot \text{OPT} \leq T \leq \text{OPT}$, OPT is the maximum **fraction** of constraints that can satisfied
  - outputs an "underestimate" that is not off by a factor more than α
  - Randomized algorithm: outputs such an estimate with probability at least $2/3$

Promise problems

- For $\beta < \gamma$, can you distinguish instances with OPT $\geq \gamma$ from OPT $\leq \beta$?
- Finer study of approximation:
  - $\alpha = \inf_{\text{indistinguishable } (\gamma,\beta)} \beta/\gamma$



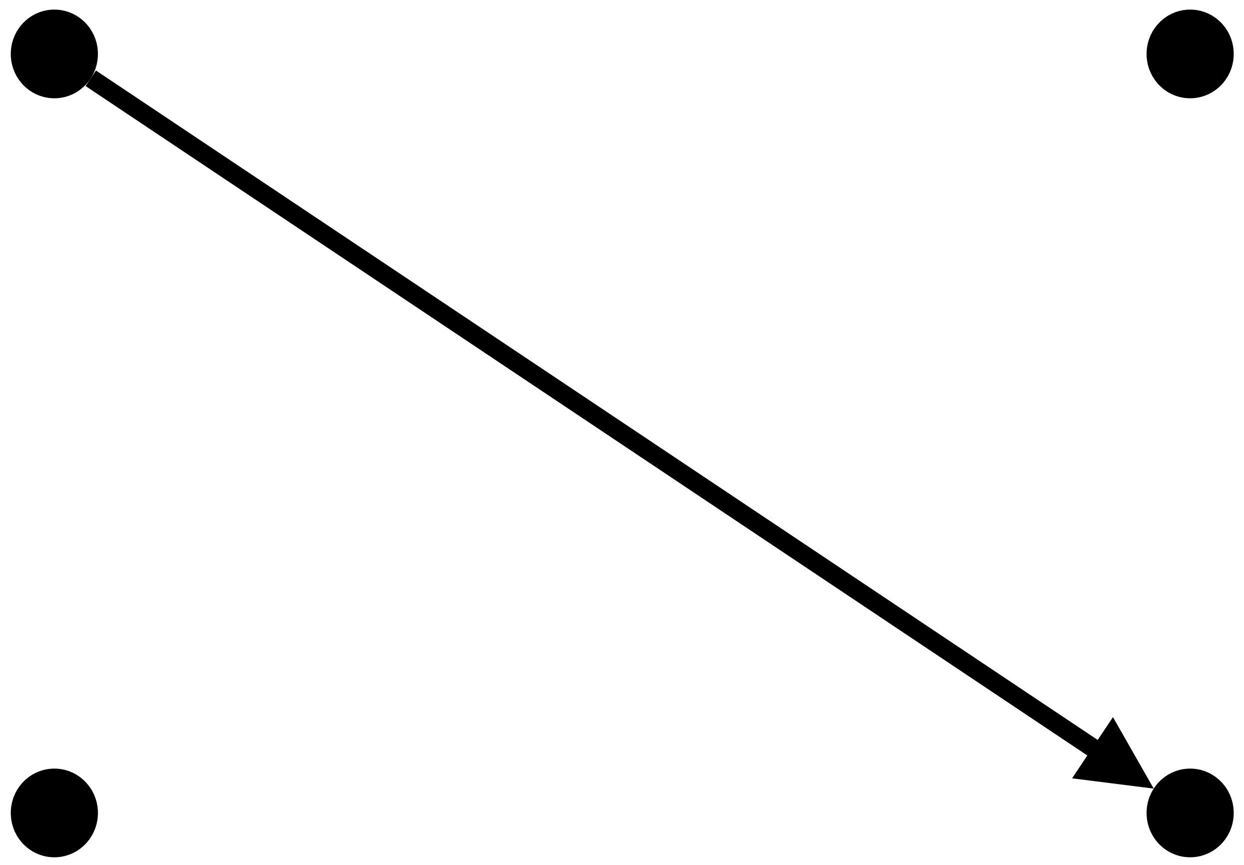
Approximability curve

Folklore approximations

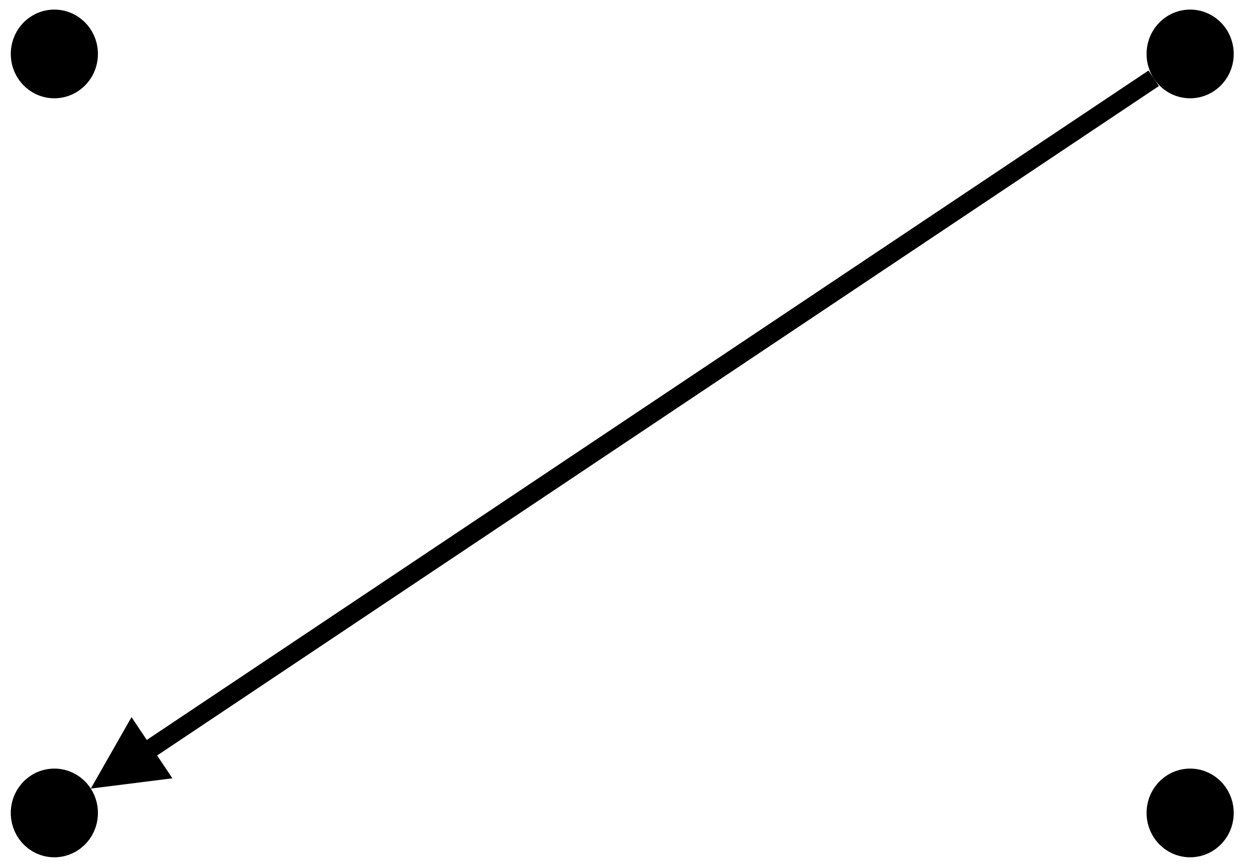
- Random CUT achieves $1/2$ approximation for Max-CUT
    - Random DICUT achieves $1/4$ approximation for Max-DICUT
- Sample $O(n/\epsilon^2)$ random edges and compute Max-CUT/Max-DICUT value (possibly in exponential time) to obtain $(1-\epsilon)$ approximation!

Streaming setting

Streaming setting



Streaming setting

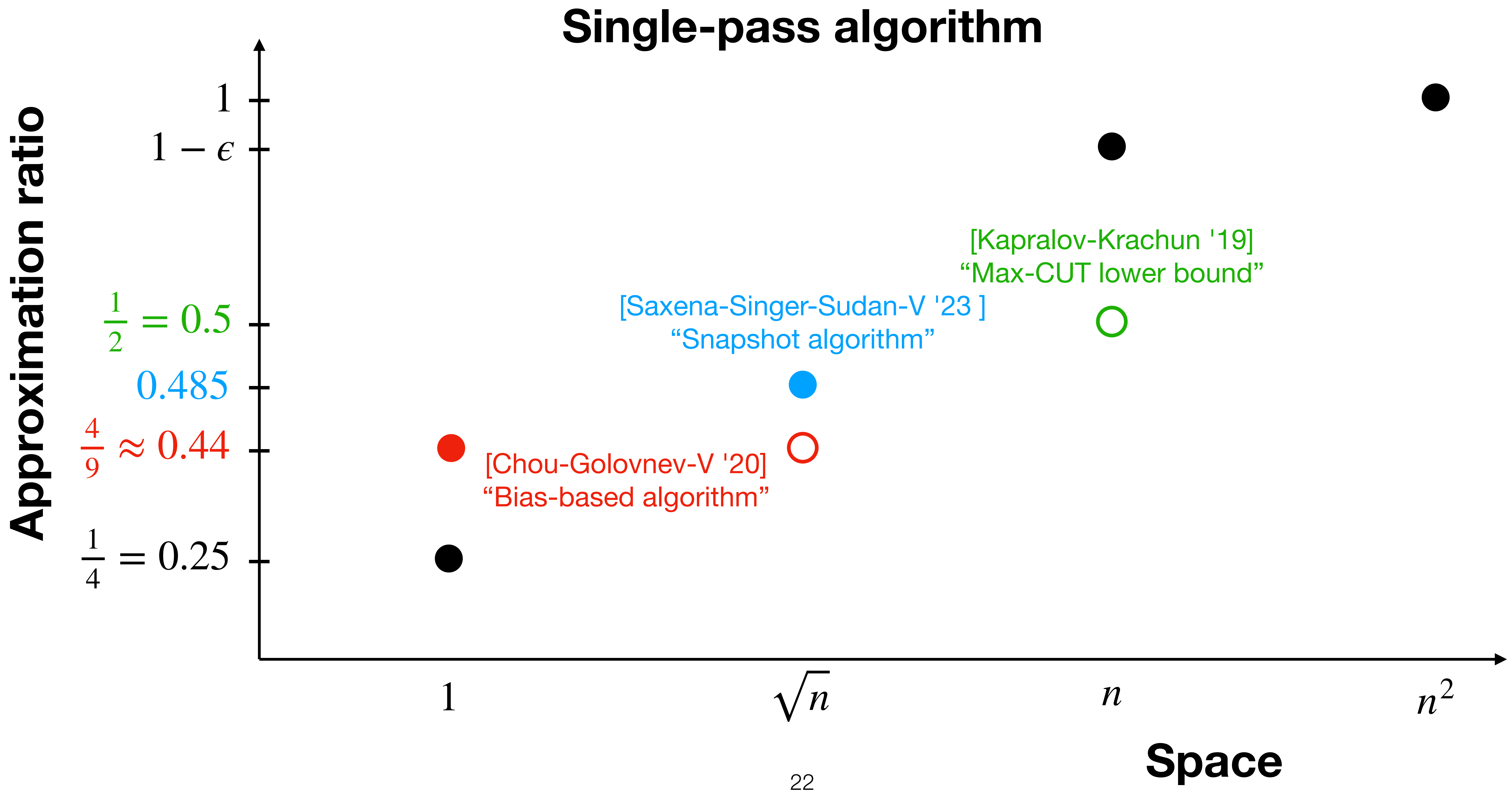


Streaming setting

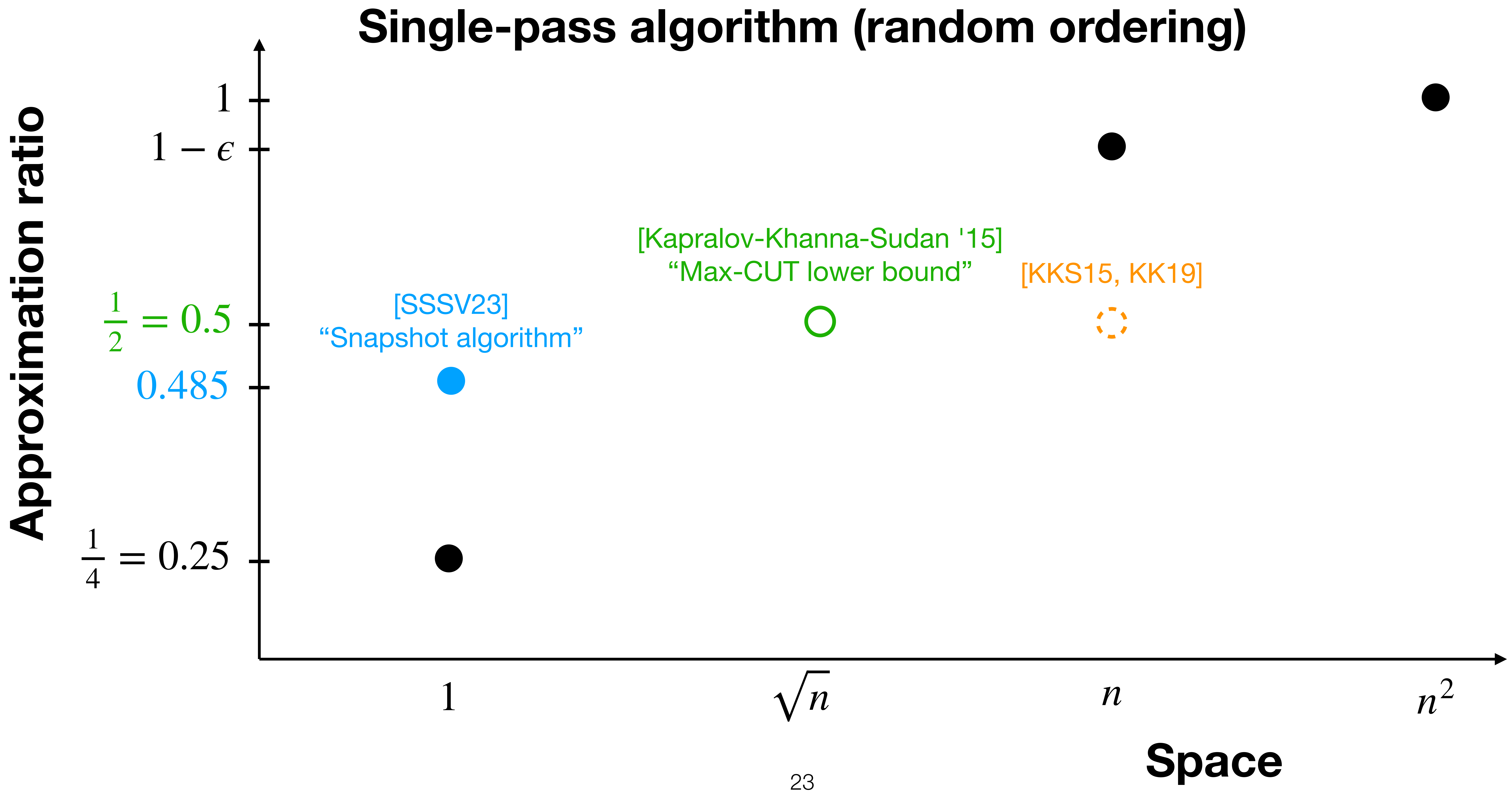
Streaming setting

Compute the size of the largest CUT/DICUT

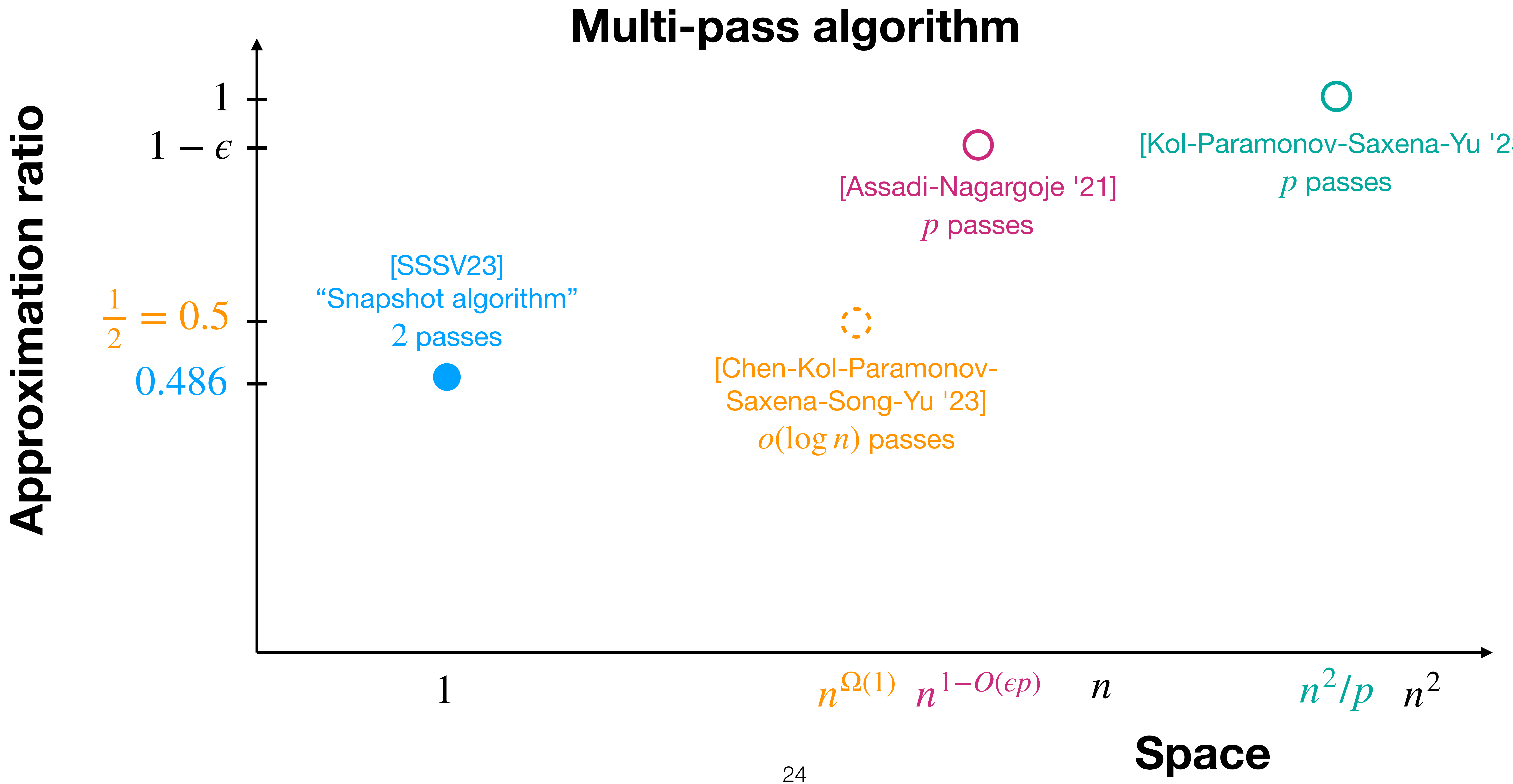
State of the art: Max-DICUT



State of the art: Max-DICUT



State of the art: Max-DICUT



Rest of the talk

- Log space sketching algorithm for Max-DICUT
- $\Omega(\sqrt{n})$ space lower bounds for Max-DICUT
- Improved approximation in
  - $\tilde{O}(\sqrt{n})$ space
  - Log space random ordering
  - two passes
- Generalizations to arbitrary CSPs

$2/5$ approximation [Guruswami-Velingker-V '17]

- bias of vertex $b_v := \text{out}_v - \text{in}_v$
- Good DICUT
  - Positively biased $\rightarrow$ left set
  - Negatively biased $\rightarrow$ right set
- Best of random DICUT and Good DICUT — $2/5$ approximation!

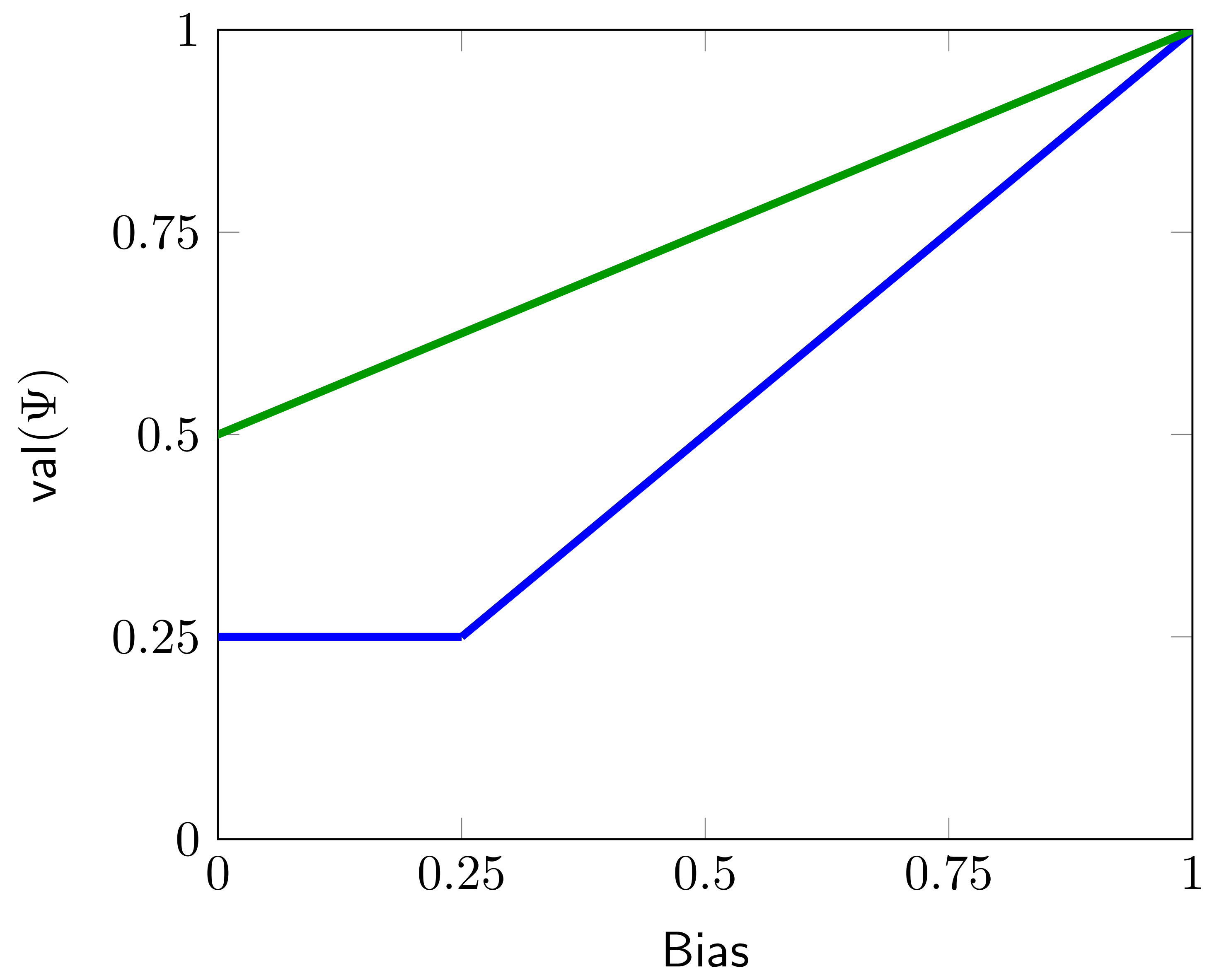
Estimation in log space

- **Bias** of the graph := $\|(b_v)\|_1$

$$\max\left\{\frac{m}{4}, \frac{\text{Bias}}{2}\right\} \leq \text{Max-DICUT} \leq \frac{m}{2} + \frac{\text{Bias}}{4}$$

Random DICUT → $\frac{m}{4}$

Good DICUT → $\frac{\text{Bias}}{2}$



- [Indyk '06] Log space sketch for ℓ_1 norm estimation under bounded constant updates
  - Edge (u, v) $\quad b_u \to b_u + 1 \quad\quad b_v \to b_v - 1$

Upper bound

For any optimal DICUT

- Every vertex has at most $\max\{\text{out}_v, \text{in}_v\}$ edges in the DICUT incident on it.
- Total #edges in DICUT $\leq \frac{1}{2}\sum_i \max\{\text{out}_v, \text{in}_v\} = (2m + \text{Bias})/4$

$$\text{Bias} = \sum_v |\text{out}_v - \text{in}_v| = \sum_v \left(\max\{\text{out}_v, \text{in}_v\} - \min\{\text{out}_v, \text{in}_v\}\right)$$

$$m = \frac{1}{2}\sum_v \text{deg}_v = \frac{1}{2}\sum_v \left(\text{out}_v + \text{in}_v\right) = \frac{1}{2}\sum_v \left(\max\{\text{out}_v, \text{in}_v\} + \min\{\text{out}_v, \text{in}_v\}\right)$$

Lower bound: Good DICUT $\geq$ Bias/2

- Simple argument due to Madhu Sudan

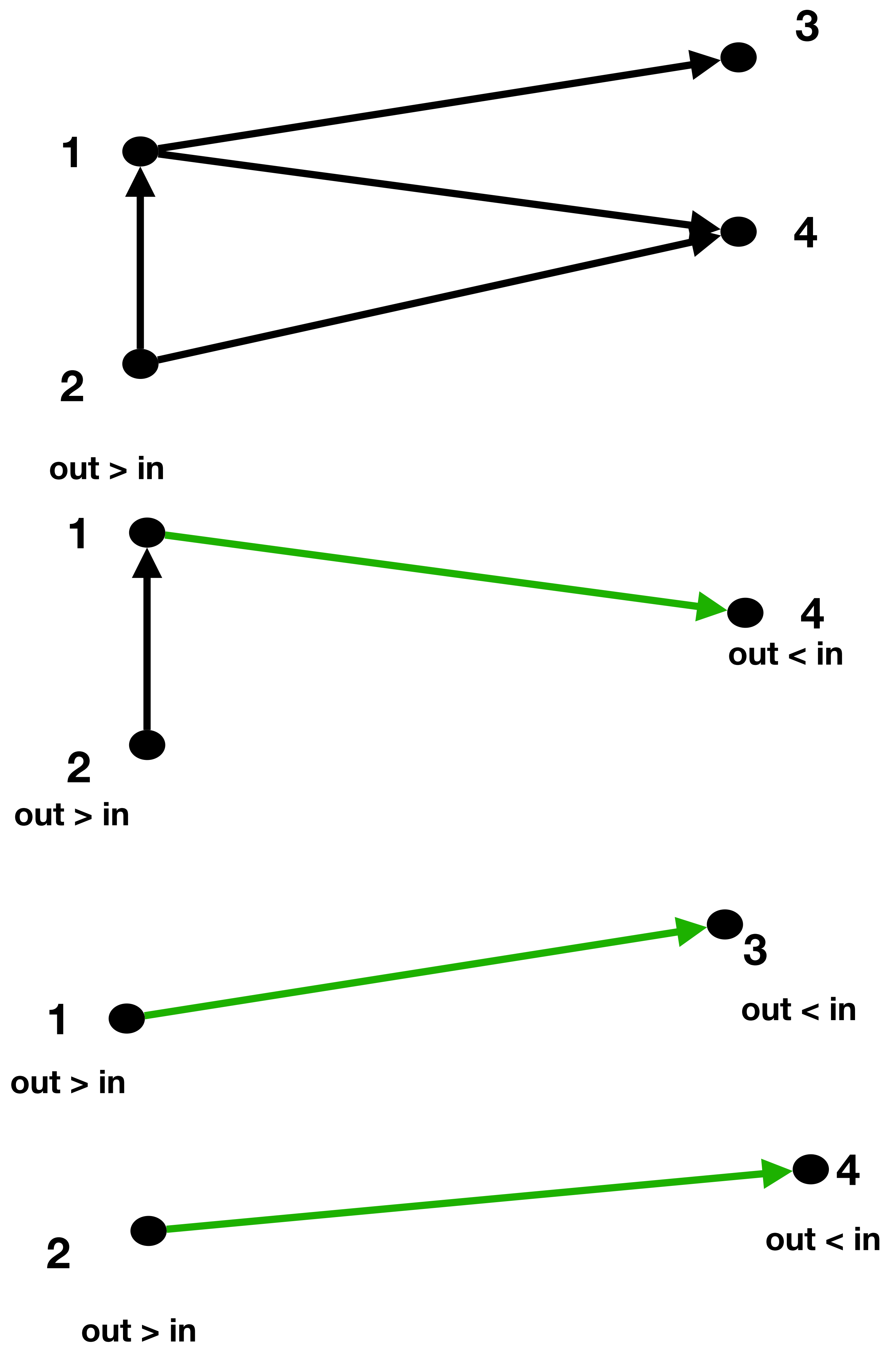
- Remove directed cycles from the graph - does not change the Bias

- In the resulting DAG, keep removing maximal paths

  - decreases Bias by 2 (the absolute value of bias of each endpoint decreases by 1, others are unchanged)

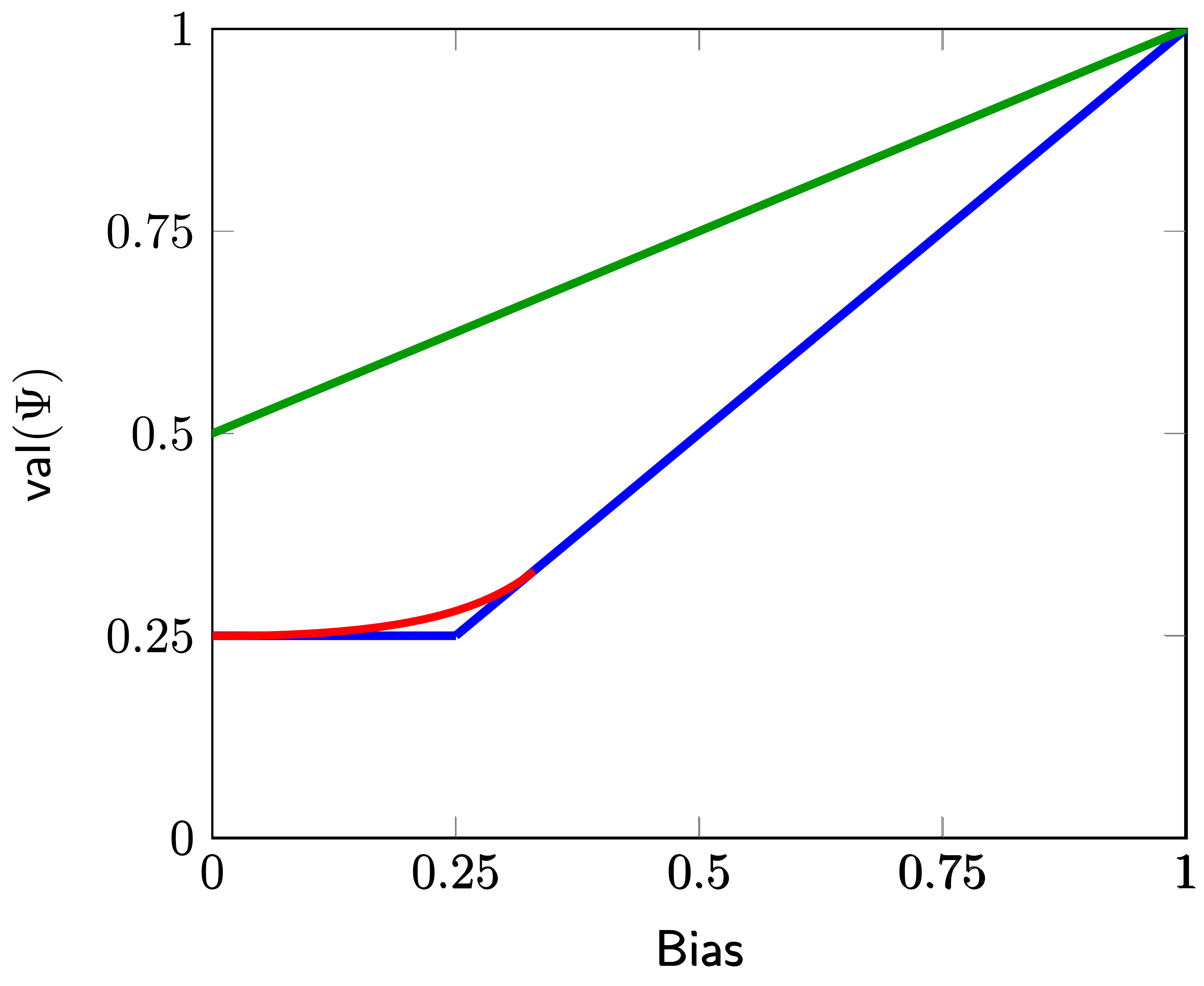
  - contributes at least 1 to the value of Good DICUT — bias of left endpoint > 0 and bias of right endpoint < 0

- Good DICUT $\geq$ Bias/2



$4/9$ approximation

- A better rounding for graphs with small Bias
  - Positively biased $\rightarrow$ left set with prob. $1/2 + \delta$
  - Negatively biased $\rightarrow$ right set with prob. $1/2 + \delta$
- $\delta = \dfrac{\text{Bias}}{2(1 - 2 \cdot \text{Bias})}$ $\rightarrow$ expected value is a quadratic function of Bias - $4/9$ approximation!
- This is also optimal in $o(\sqrt{n})$ space!

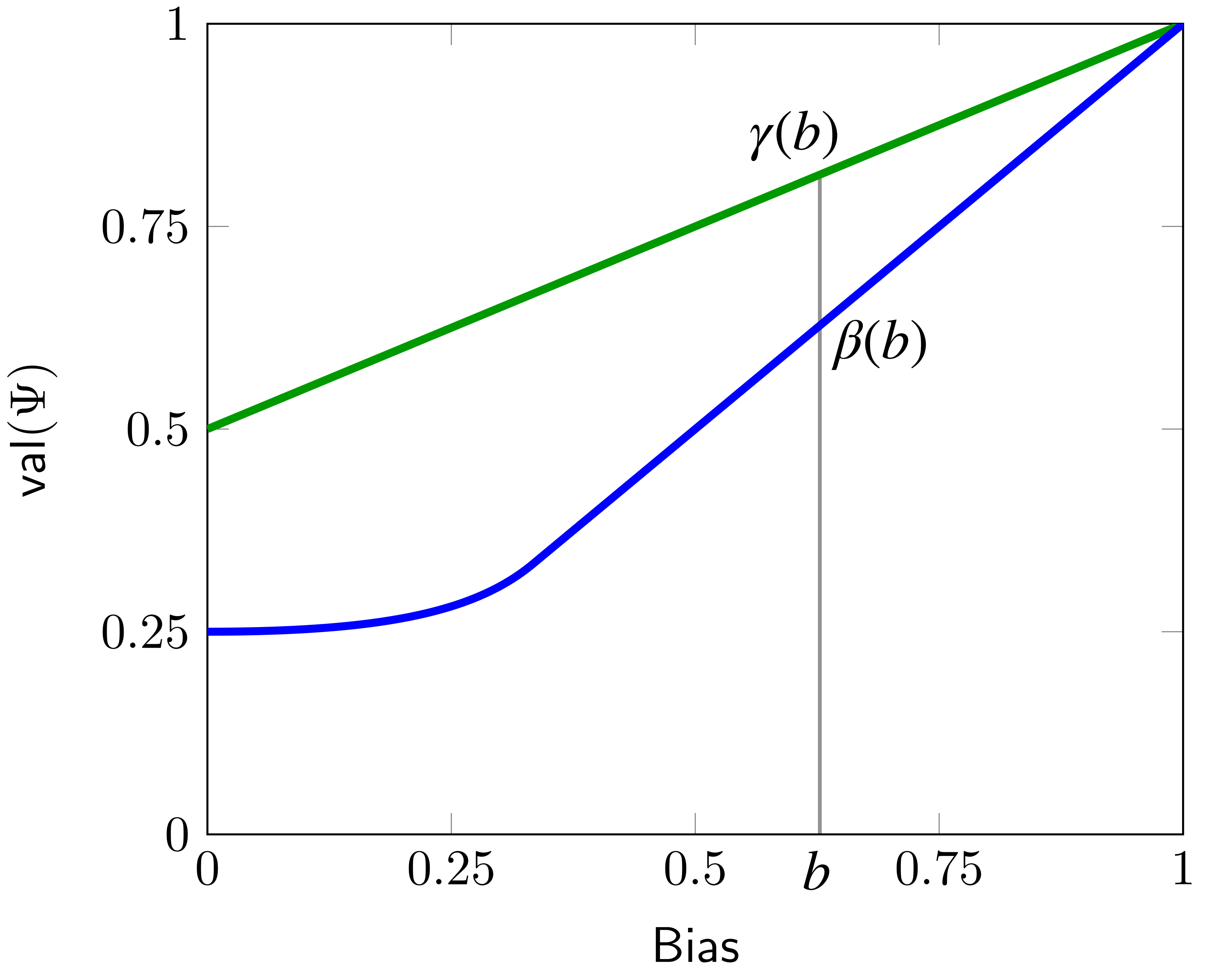


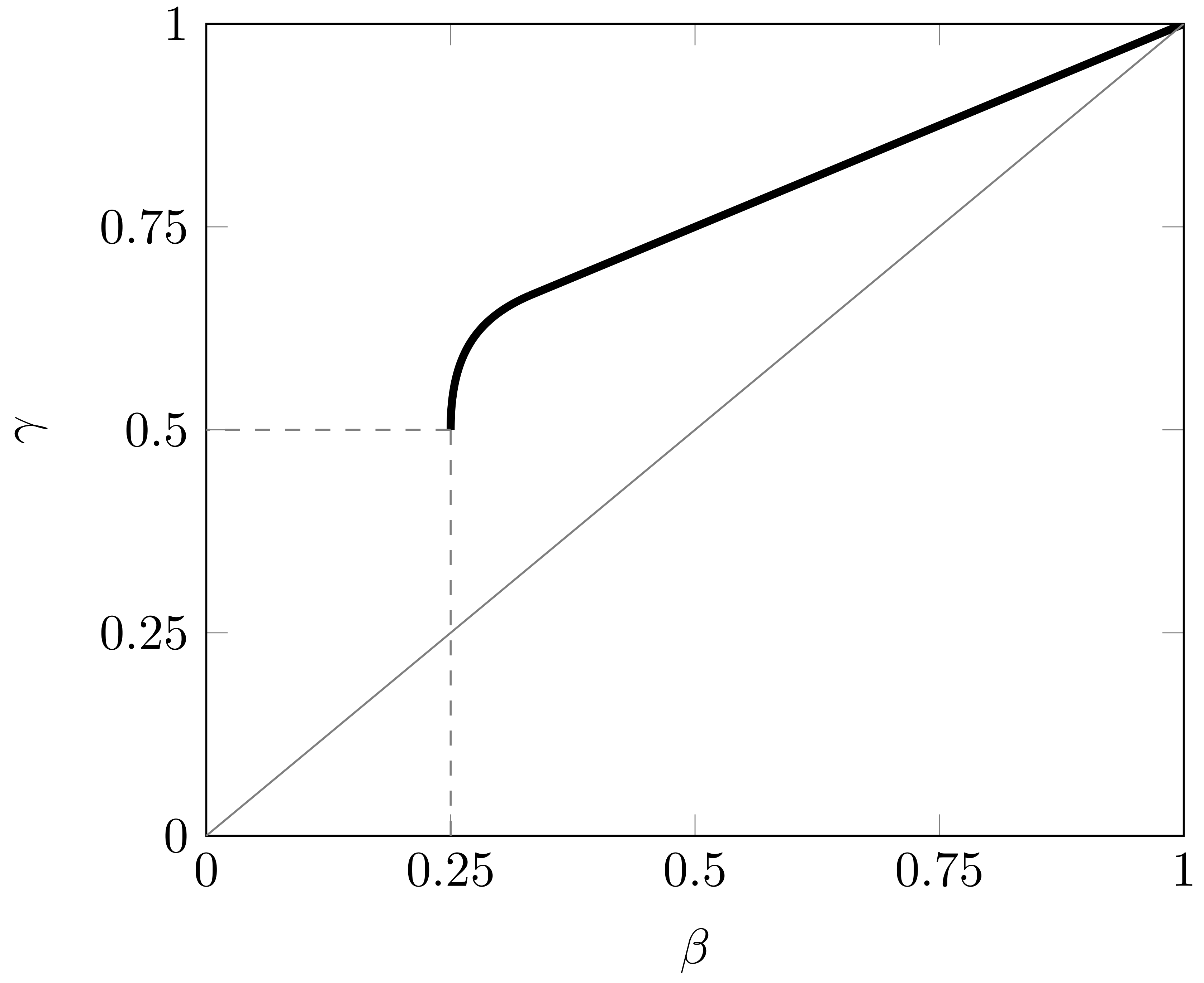
[GVV17]

Dichotomy theorem

- For $\beta < \gamma$, can you distinguish instances with OPT $\geq \gamma$ from OPT $\leq \beta$?
- For every (γ, β)
  - Either sketch their Bias to distinguish the instances
  - or any streaming algorithm that correctly distinguishes requires $\sqrt{n}$ space

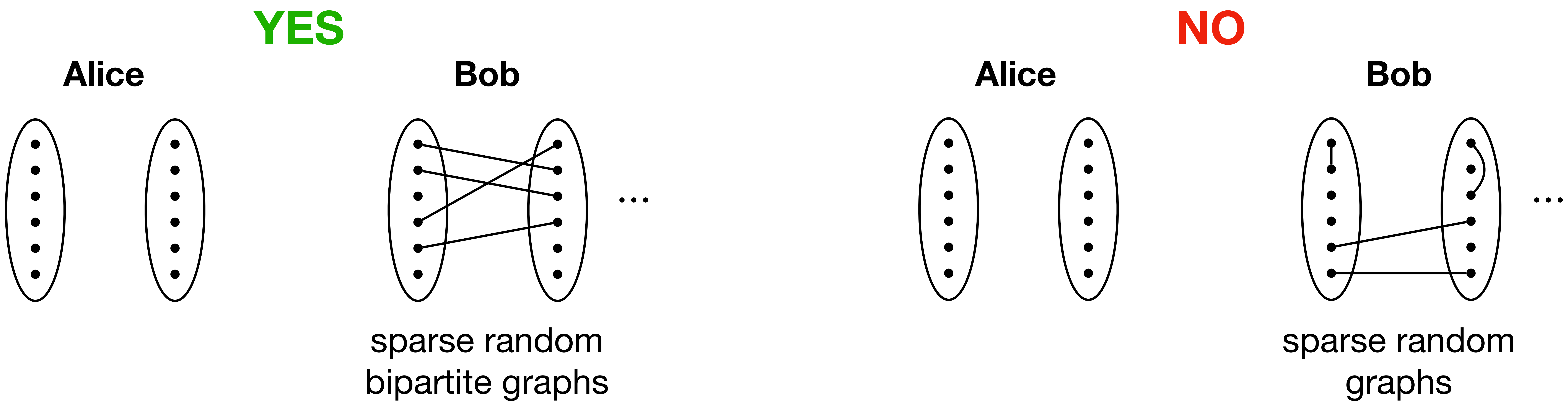
Approximability curve





$\Omega(\sqrt{n})$ space lower bounds

Distributional Boolean Hidden Partition (DBHP)

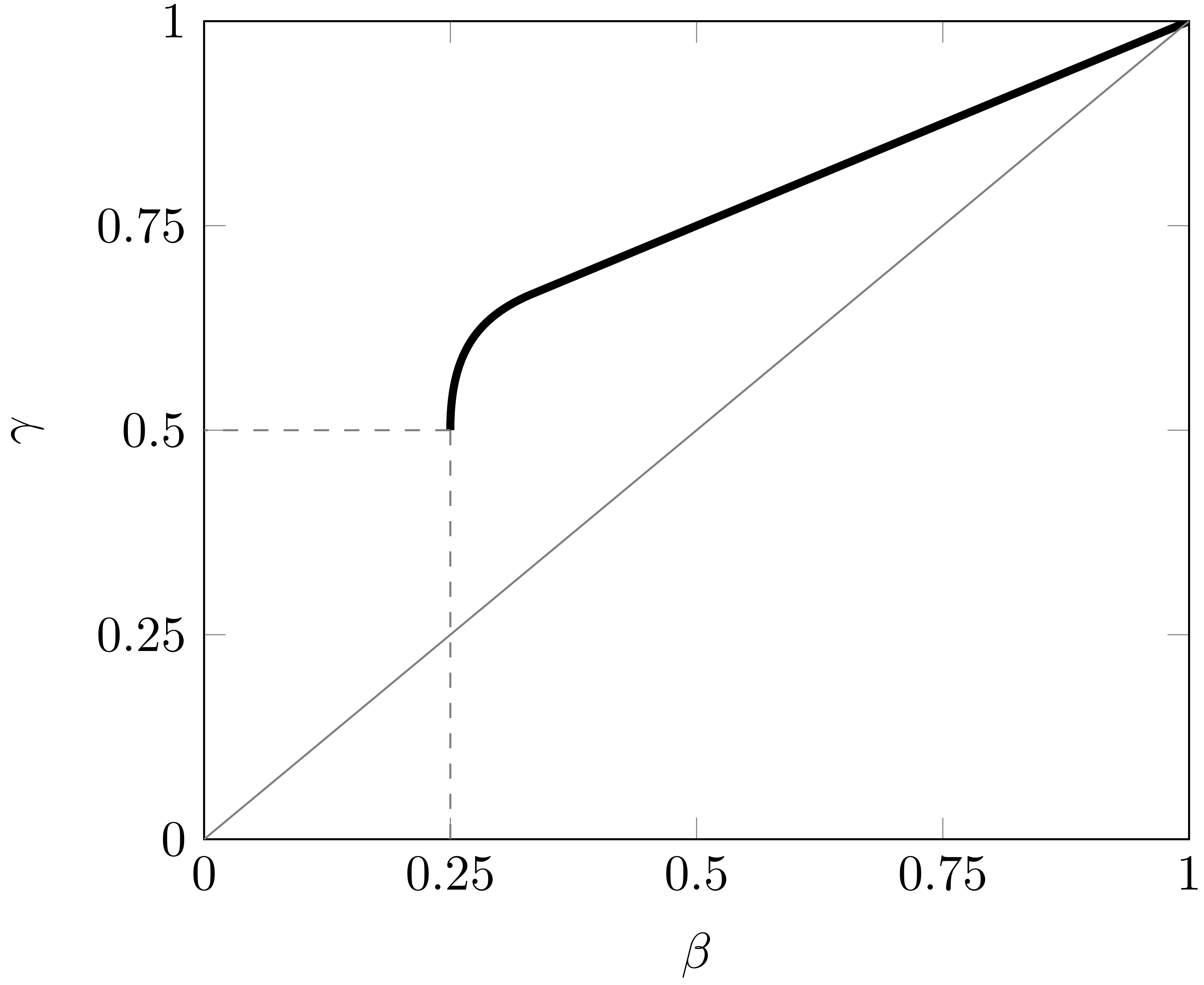


- Goal: distinguish between two cases with prob $> 2/3$
- Birthday paradox: $O(\sqrt{n})$ bits of communication suffices
- Lower bound [Gavinsky-Kempe-Kerenidis-Raz-Wolf '07, Verbin-Yu '11, Kapralov-Khanna-Sudan '15]: $\Omega(\sqrt{n})$ bits of communication is necessary

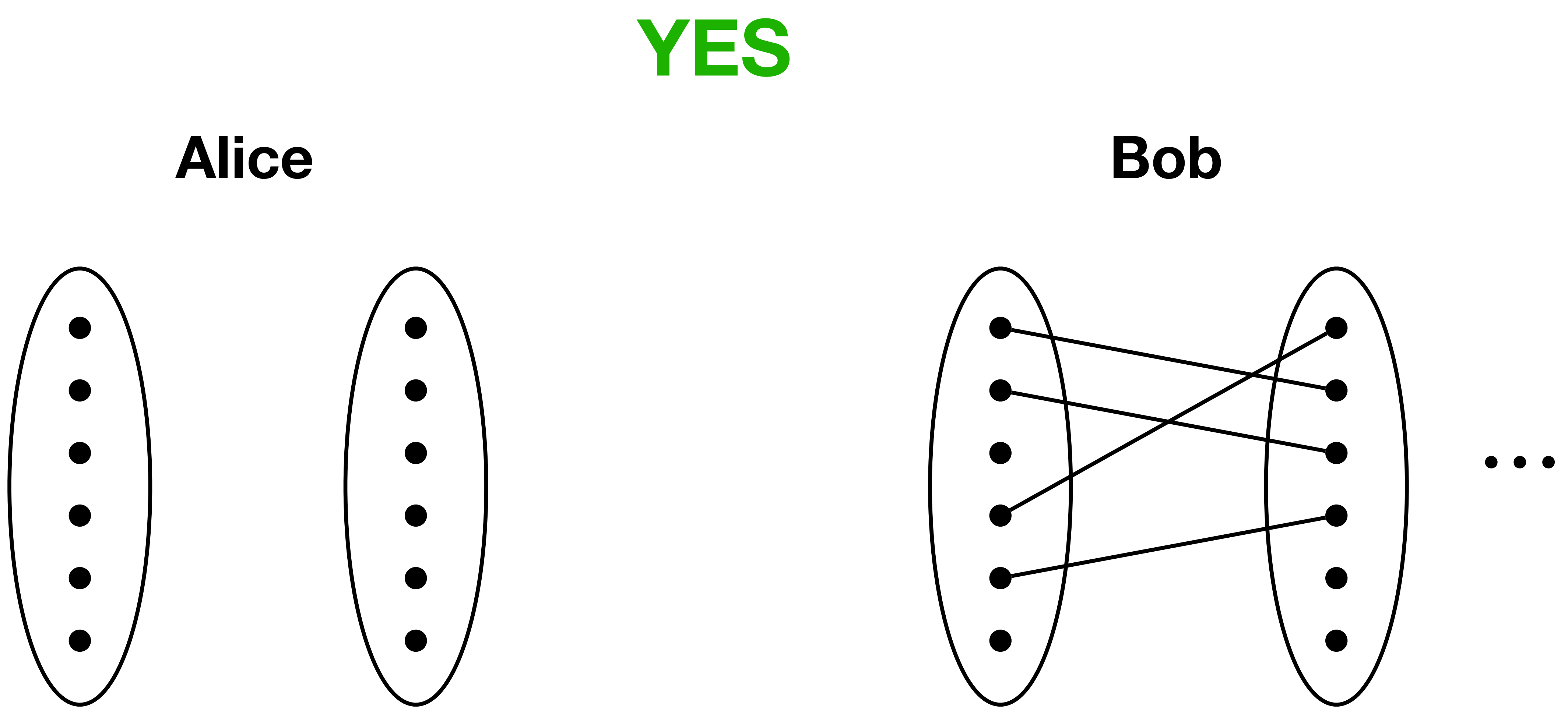
Reduction to streaming Max-DICUT

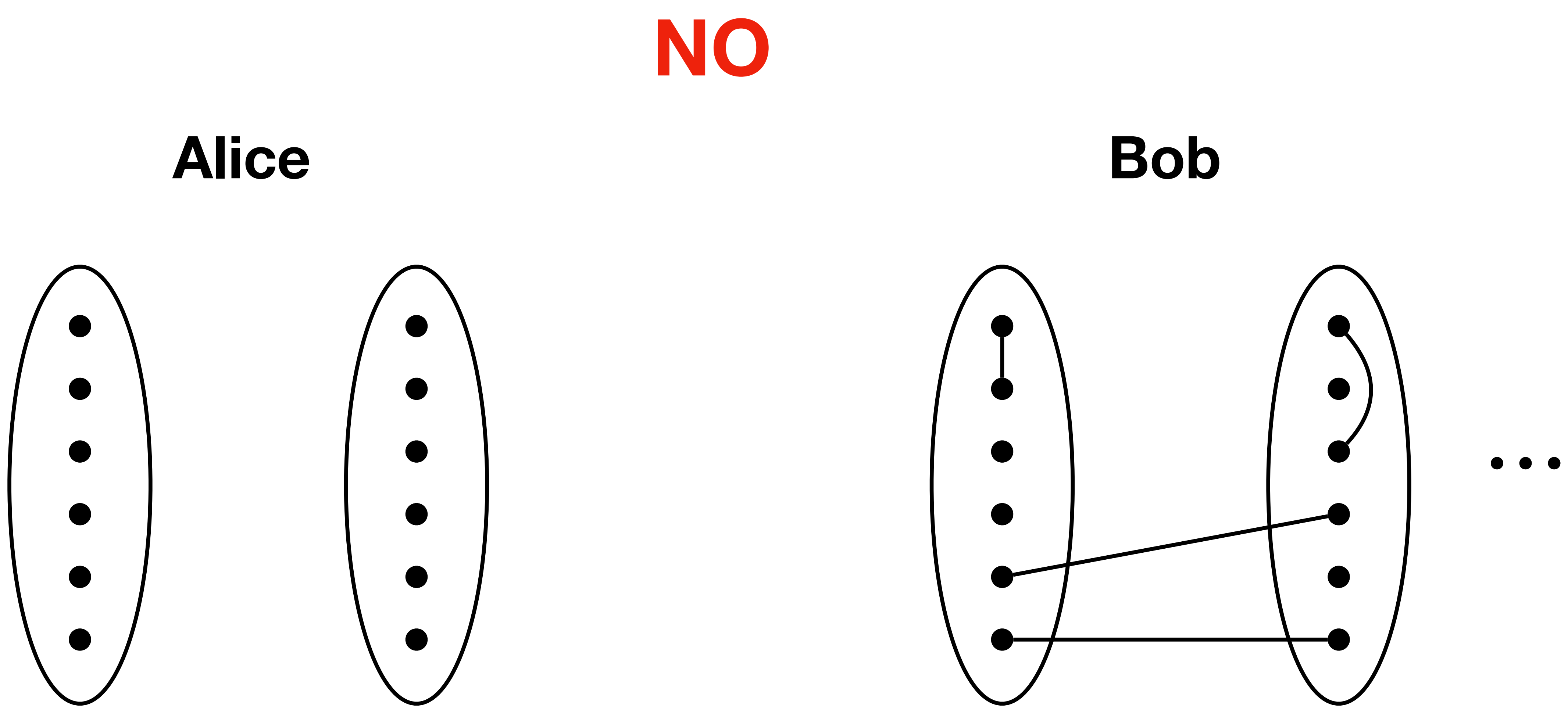
YES Max-DICUT $\geq \gamma - \epsilon$

NO Max-DICUT $\leq \beta + \epsilon$

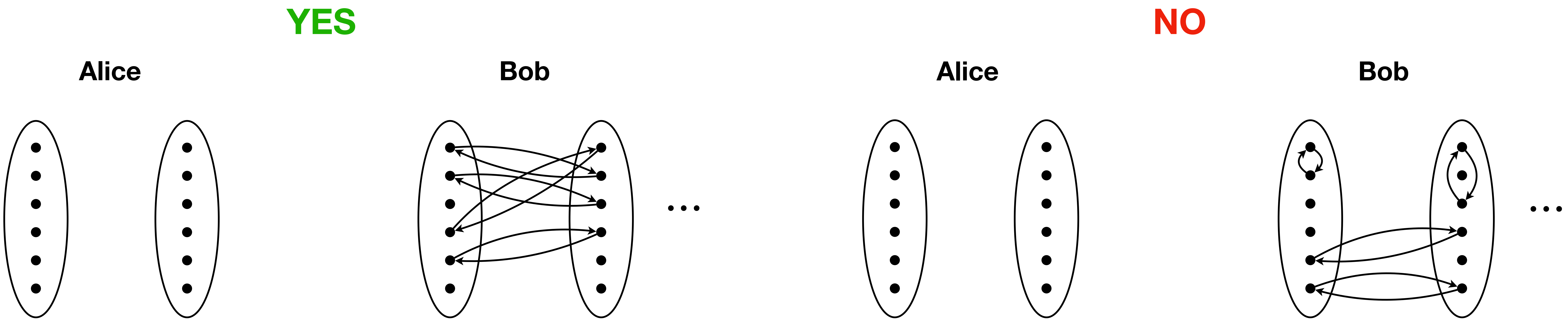


Reduction to streaming Max-DICUT





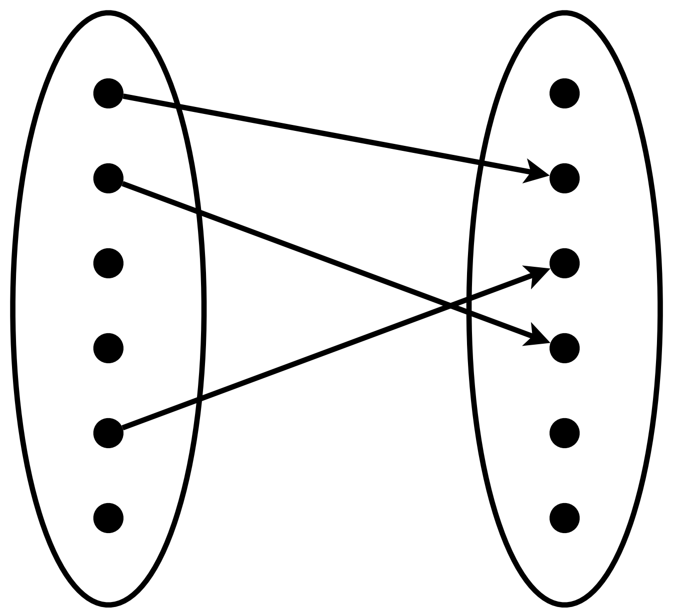
Reduction to streaming Max-DICUT



Reduction to streaming Max-DICUT

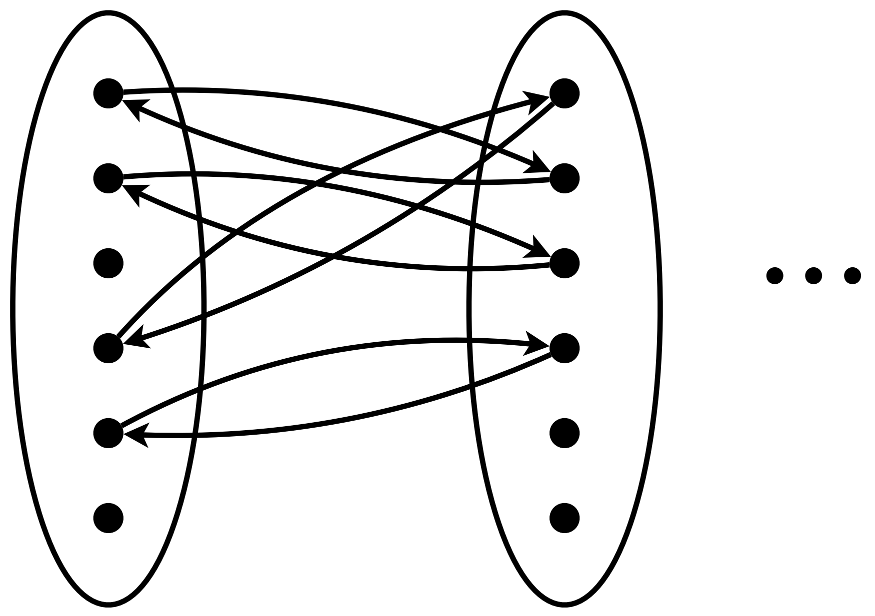
YES

Alice



$\frac{bM}{1-b}$ edges

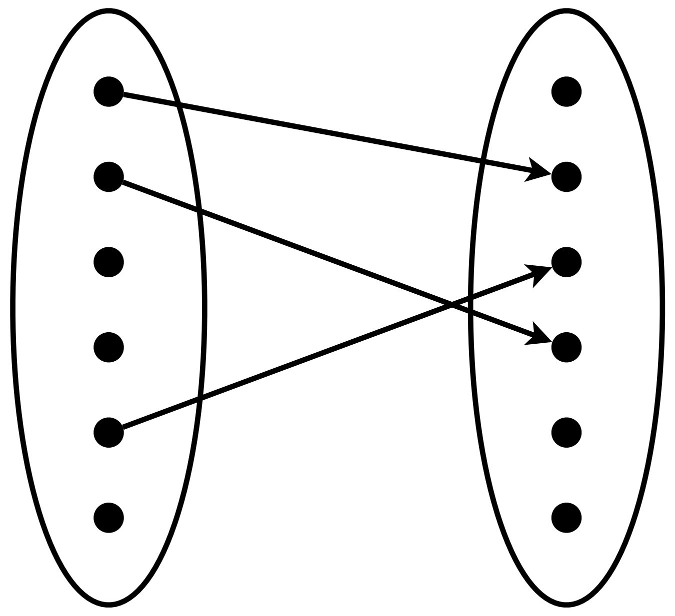
Bob



M edges

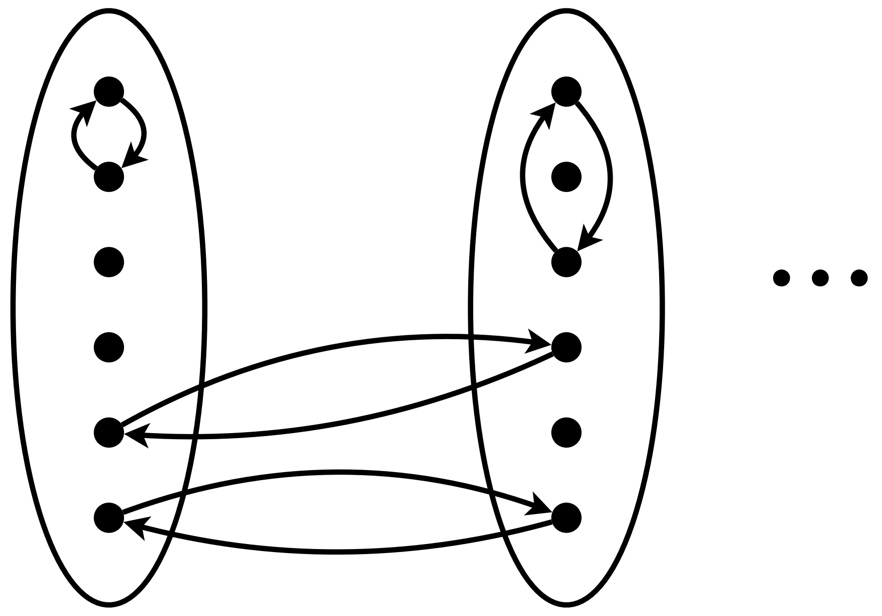
NO

Alice



$\frac{bM}{1-b}$ edges

Bob

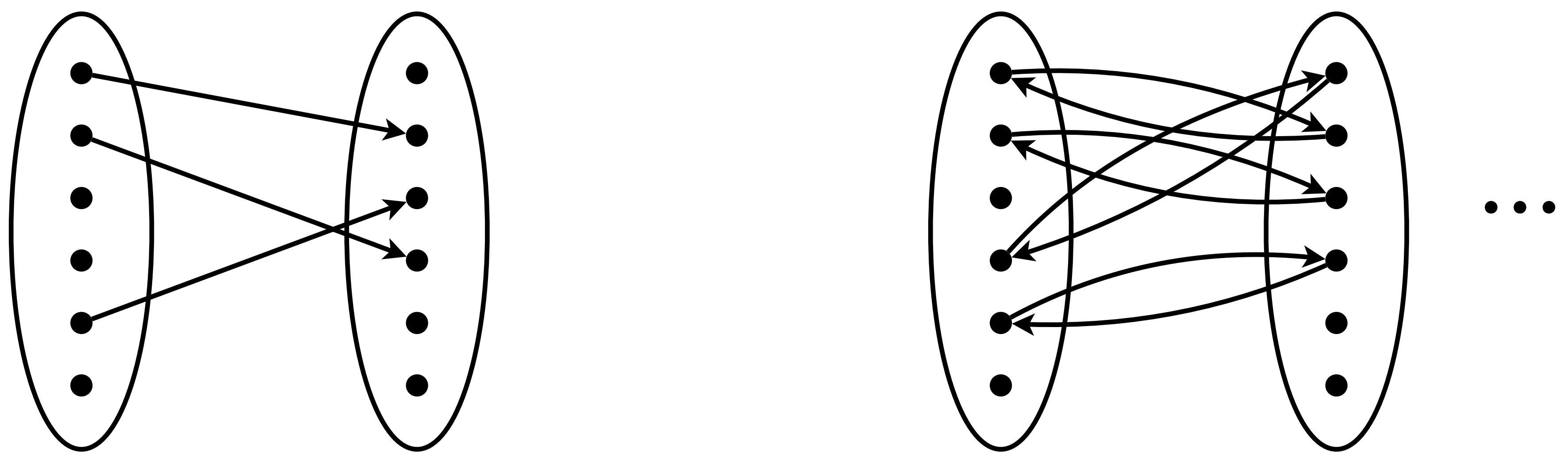


M edges

Reduction to streaming Max-DICUT

YES

Alice **Bob**

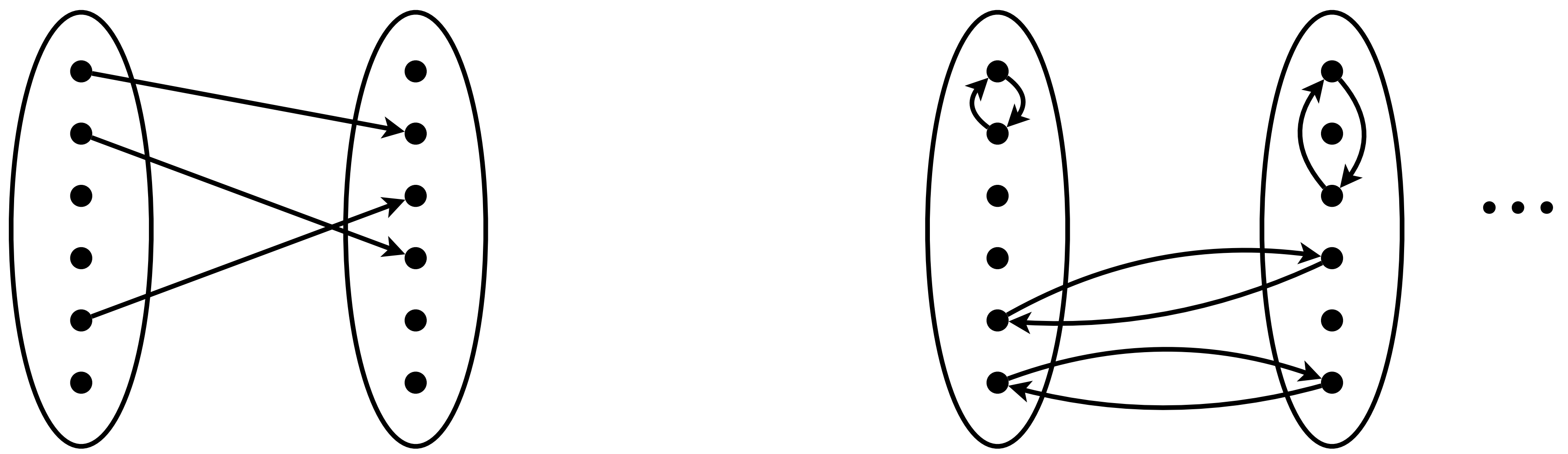


$\frac{bM}{1-b}$ edges

M edges

NO

Alice **Bob**

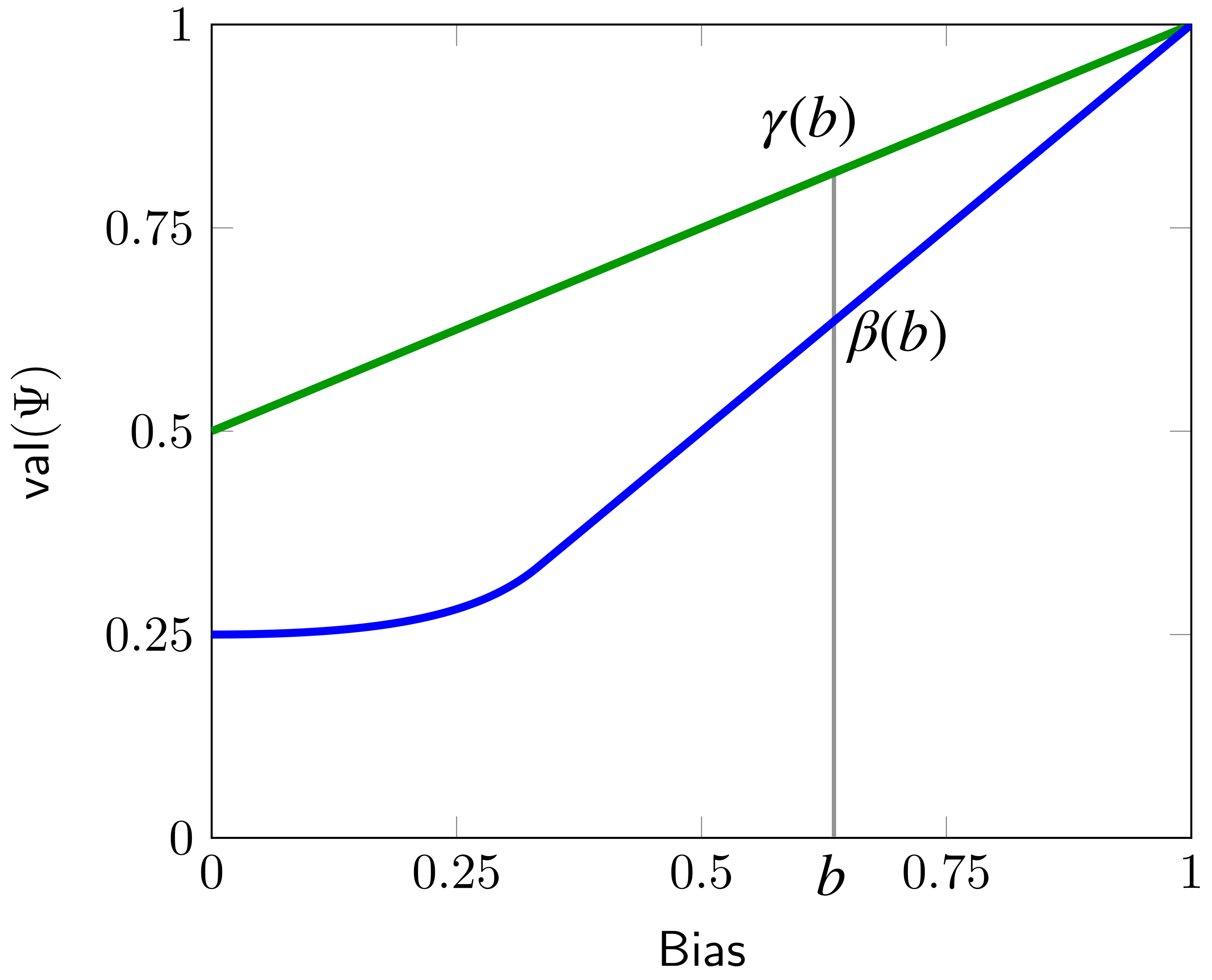


$\frac{bM}{1-b}$ edges

M edges

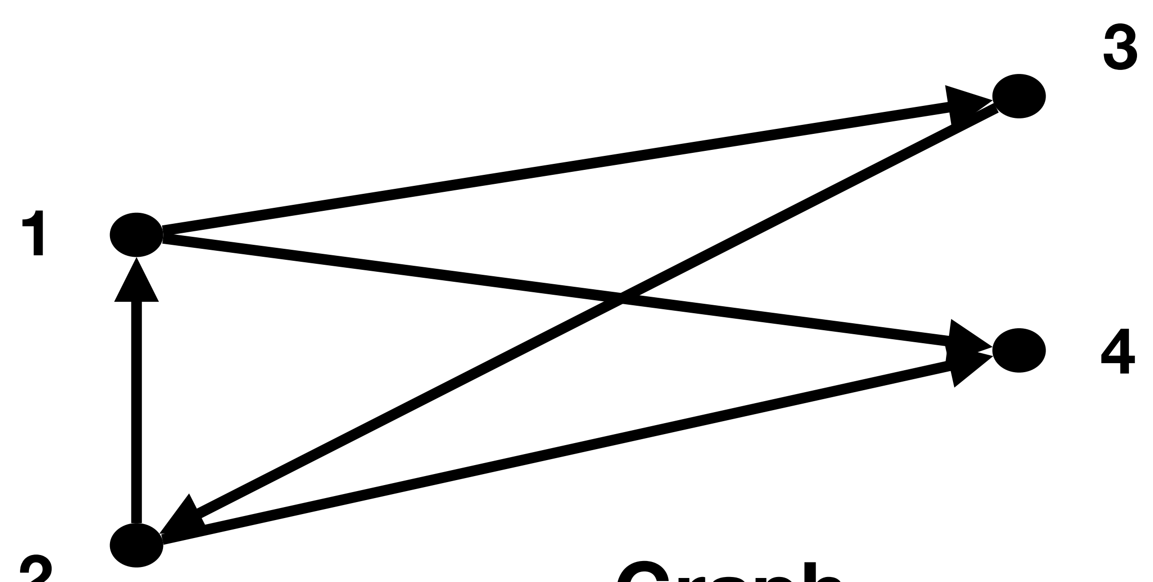
YES Max-DICUT $\geq \gamma(b) - \epsilon$

NO Max-DICUT $\leq \beta(b) + \epsilon$

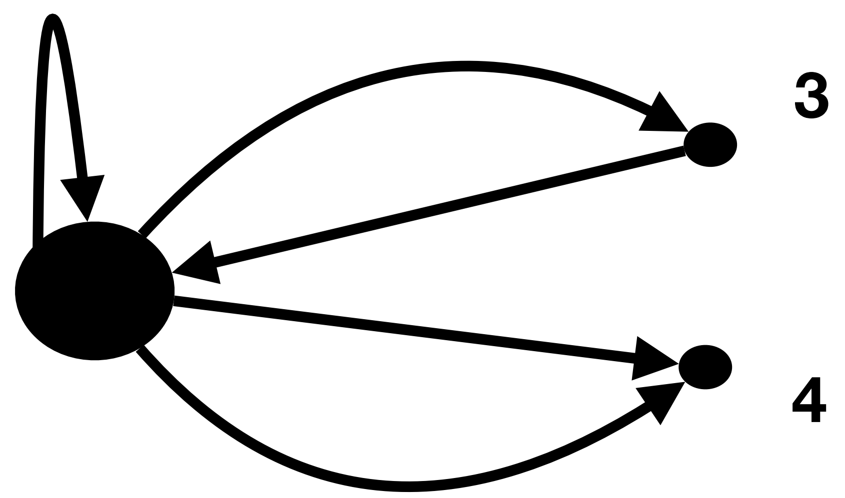


Beating $4/9$ approximation in $\tilde{O}(\sqrt{n})$ space

- Sketch "snapshot" of the graph in $\tilde{O}(\sqrt{n})$ space
- Use the snapshot to get better approximation



Graph

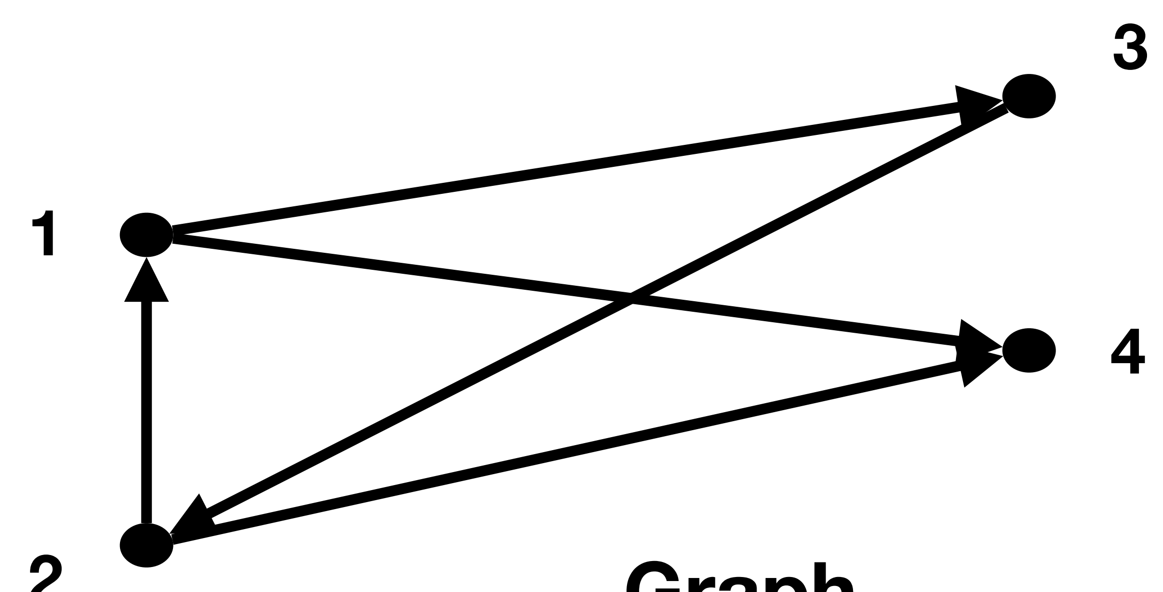


Snapshot

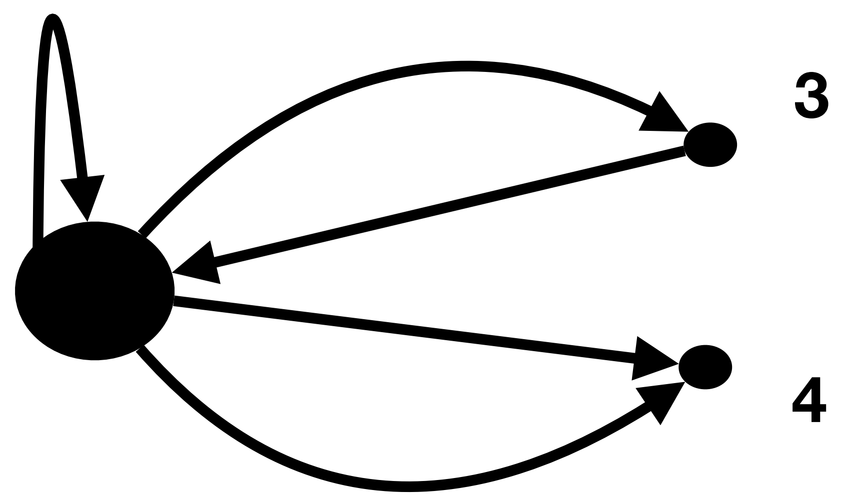
Collapse vertices of same bias

Beating $4/9$ approximation in $\tilde{O}(\sqrt{n})$ space

- Sketch "snapshot" of the graph in $\tilde{O}(\sqrt{n})$ space
- **Use the snapshot to get better approximation**



Graph



Snapshot

Collapse vertices of same bias

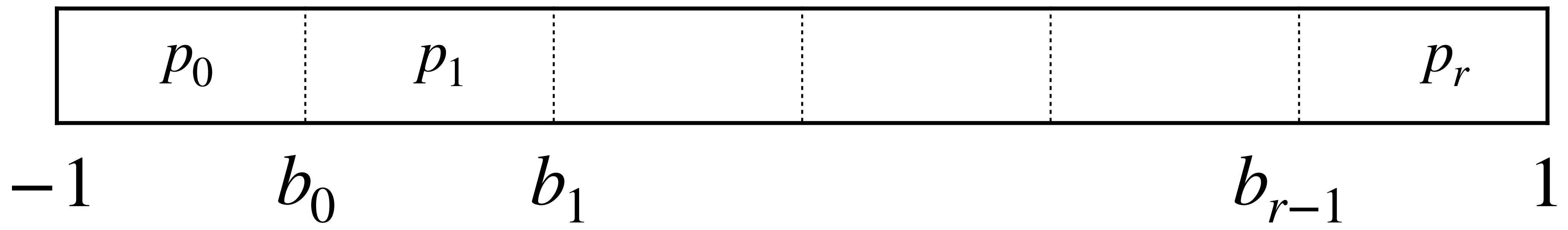
Oblivious-algorithms for Max-DICUT

[Feige-Jozeph '15]

- Indifferent between vertices of same bias ≡ Snapshot is the input to the algorithm

- Selection function: Bias-intervals

$-1 < b_0 < b_1 < \cdots < b_{r-1} < 1$



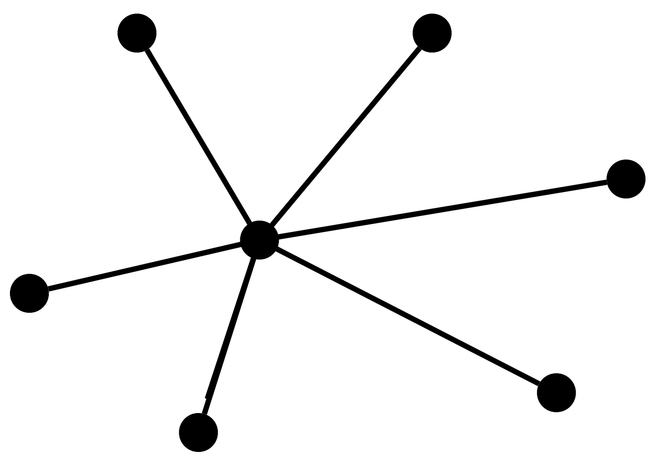
- Selection function with $0.483 > 4/9$ approximation!

- Lower bound: Graphs with same snapshot but Max-DICUT value ratio 0.4899

- Improvements in third decimal place [Hwang-Singer-V '23]

Snapshot estimation

- WLOG assume that the graph is sparse
- Random ordering: Take the first $O(1)$ edges. Store the induced subgraph and track the biases of the vertices. Compute snapshot.
- Adversarial ordering + Two passes: Use the first pass to sample $O(1)$ random edges.
- Bounded-degree graphs: Sample $O(\sqrt{n})$ random non-isolated vertices. Store the induced subgraph and track the biases. Compute snapshot.
  - Bounded-degree assumption is crucial!



Snapshot estimation for arbitrary graphs

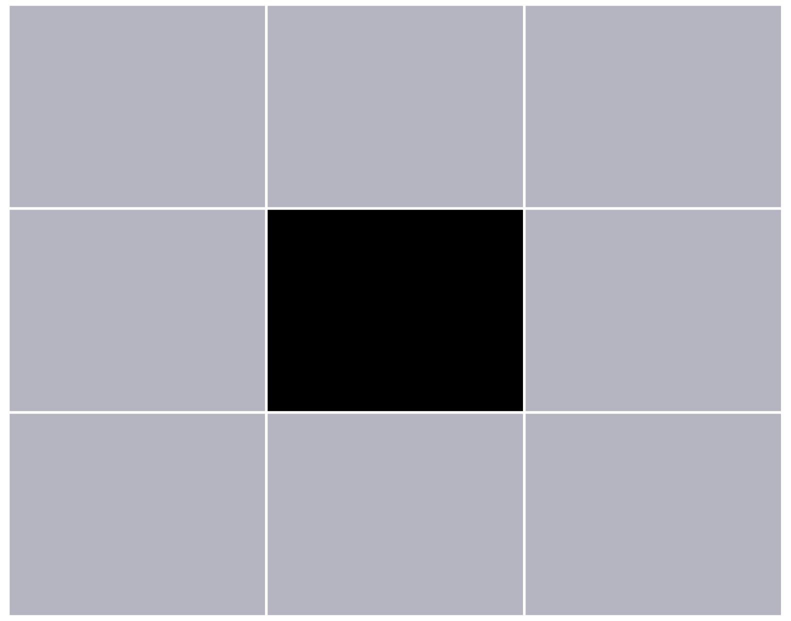
- "Layered Sampling of vertices"
  - Partition the degrees: $1 \leq d_0 < d_1 \cdots < d_{\log n} \leq n$, where $d_{i+1}/d_i = 2$
  - Obtain a uniform sample of vertices in each degree-interval and store the induced subgraph
    - sample vertices of degree d with probability $\min\left\{\tilde{O}\left(\frac{d}{\sqrt{n}}\right), 1\right\}$
- Issue: We do not know the degree of the vertex when it first appears in the stream

Snapshot estimation for arbitrary graphs

- "Layered sampling of edges"
  - Subsample edges with probability $\tilde{O}\left(\frac{1}{d}\right)$
  - Degree d vertices have degree $\tilde{O}(1)$ in the above graph
  - Subsample every vertex with probability $\min\left\{\tilde{O}\left(\frac{d}{\sqrt{n}}\right), 1\right\}$ and ignore vertices with degree $\gg \tilde{O}(1)$ or $\ll \tilde{O}(1)$; store all the edges incident to it
- Issue: Can misclassify vertices and place them in wrong layers. Could also potentially make errors in bias computation

Snapshot estimation for arbitrary graphs

- “Smoothed estimates of refined snapshots”
  - Refined snapshot $M((i,j),(k,l))$: How many edges go from bias-interval B_i, degree-interval D_j to bias-interval B_k and degree-interval D_l?
  - Smoothed snapshot $\widehat{M}$: Average of M over a “window” of size w
  - Pointwise estimate $\widehat{M}$. Better estimates with increasing w



Why smoothed snapshots suffice?

Why don't misclassification errors matter?

- Degree misclassification errors do not affect snapshot computation
  - If there is no error in bias computation, snapshot computed from M and $\widehat{M}$ are exactly the same
- Bias misclassification errors do not affect Max-DICUT value when w is small
  - Perturb the original graph G to get a new graph H with similar Max-DICUT value; Smoothed-snapshot$(G) \approx$ Refined-snapshot(H)
  - If bias of v is off by ϵ, create a new isolated vertex and modify bias by creating at most ϵd new edges; Max-DICUT value does not change by more than ϵm

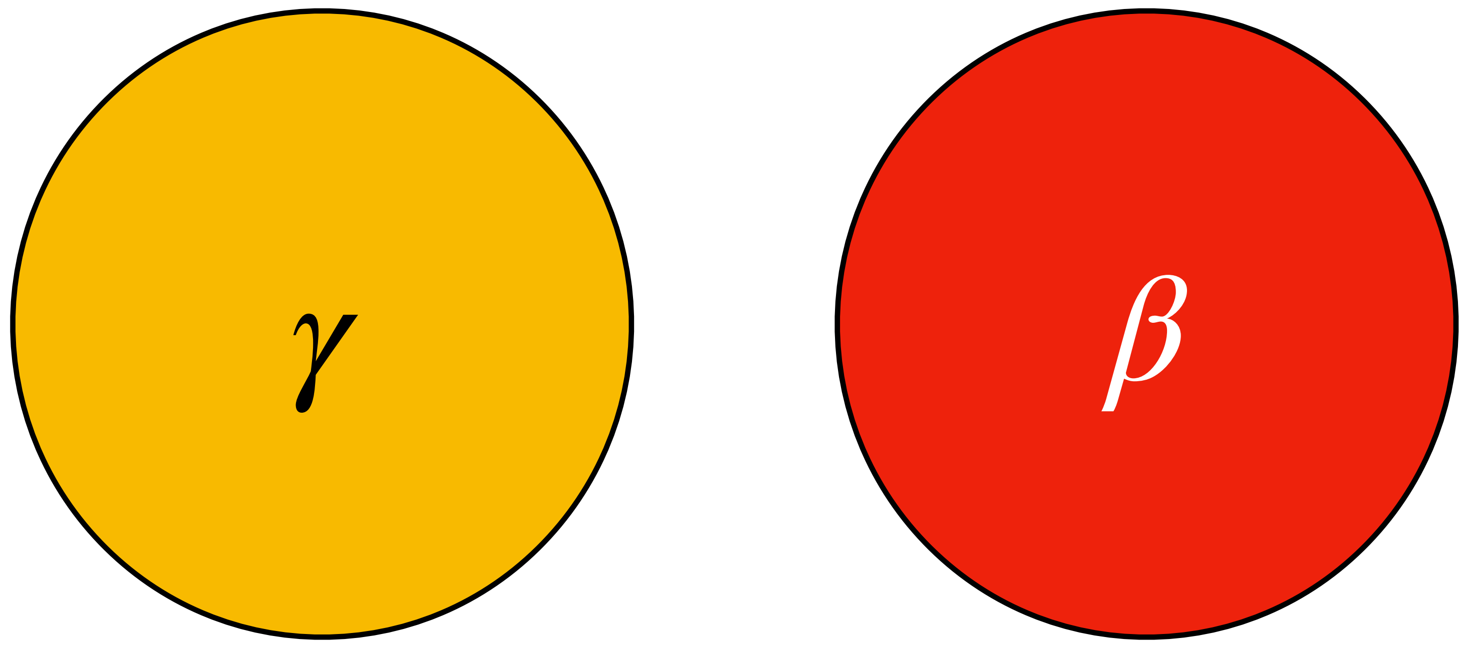
Generalizations to arbitrary CSPs

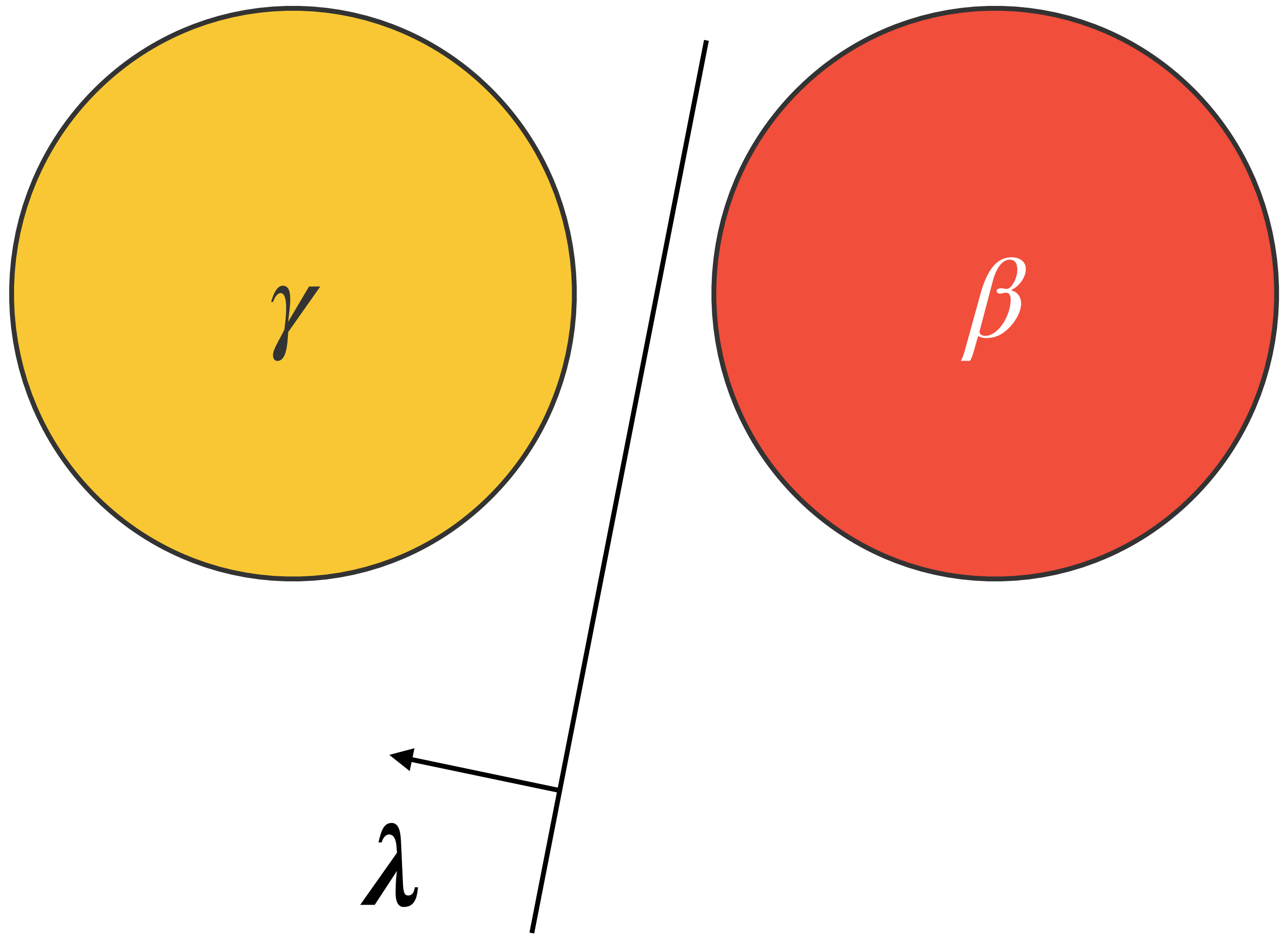
Dichotomy theorem for arbitrary CSPs

- For $\beta < \gamma$, can you distinguish instances with OPT $\geq \gamma$ from OPT $\leq \beta$?
- [Chou-Golovnev-Sudan-V '21] Dichotomy theorem for general CSPs: For every (γ, β)
  - Either sketch Bias to distinguish the instances
  - or any sketching algorithm that correctly distinguishes requires $\sqrt{n}$ space

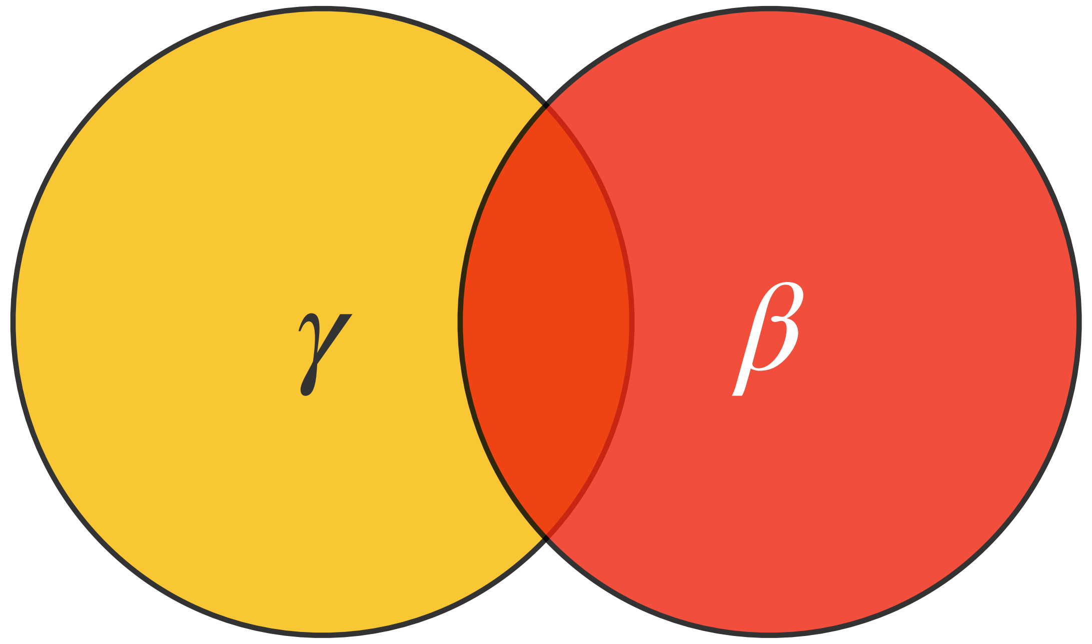
Dichotomy theorem

Convex, closed, and bounded sets in finite-dimensional space





"Bias" can distinguish



Cannot distinguish in $o(\sqrt{n})$ space

Generalized Bias

$$\text{Bias} = \max_{\mathbf{a}=(a_1,\ldots,a_n)\in\{-1,1\}^n} [1\ -1] \begin{bmatrix} \text{out}_1 & \cdots & \text{out}_n \\ \text{in}_1 & \cdots & \text{in}_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Generalized Bias

$$\text{Bias} = \max_{\mathbf{a}=(a_1,\ldots,a_n)\in\{-1,1\}^n} \underbrace{[1 \;\; -1]}_{\lambda} \begin{bmatrix} \text{out}_1 & \cdots & \text{out}_n \\ \text{in}_1 & \cdots & \text{in}_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Generalized Bias

$$\text{Bias} = \max_{\mathbf{a}=(a_1,\ldots,a_n)\in\{-1,1\}^n} \underbrace{\begin{bmatrix}1 & -1\end{bmatrix}}_{\lambda} \underbrace{\begin{bmatrix}\text{out}_1 & \cdots & \text{out}_n \\ \text{in}_1 & \cdots & \text{in}_n\end{bmatrix}}_{\mathbf{B}} \begin{bmatrix}a_1 \\ \vdots \\ a_n\end{bmatrix}$$

Generalized Bias

$$\text{Bias} = \max_{\mathbf{a}=(a_1,\ldots,a_n)\in\{-1,1\}^n} \underbrace{[1\ \ -1]}_{\lambda} \underbrace{\begin{bmatrix} \text{out}_1 & \cdots & \text{out}_n \\ \text{in}_1 & \cdots & \text{in}_n \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}}_{\mathbf{a}^T} = \|\lambda B\|_1$$

Boolean CSPs: For distinguishable (γ, β),

$$\text{Bias} = \max_{\mathbf{a}\in\{-1,1\}^n} \underbrace{\boldsymbol{\lambda}_{\gamma,\beta}}_{\in\mathbb{R}^{1\times k}} \underbrace{\mathbf{B}}_{\in\mathbb{R}^{k\times n}} \mathbf{a}^T = \|\boldsymbol{\lambda}_{\gamma,\beta}\mathbf{B}\|_1$$

General CSPs: $\boldsymbol{\lambda}_{\gamma,\beta} \in \mathbb{R}^{q\times k}$

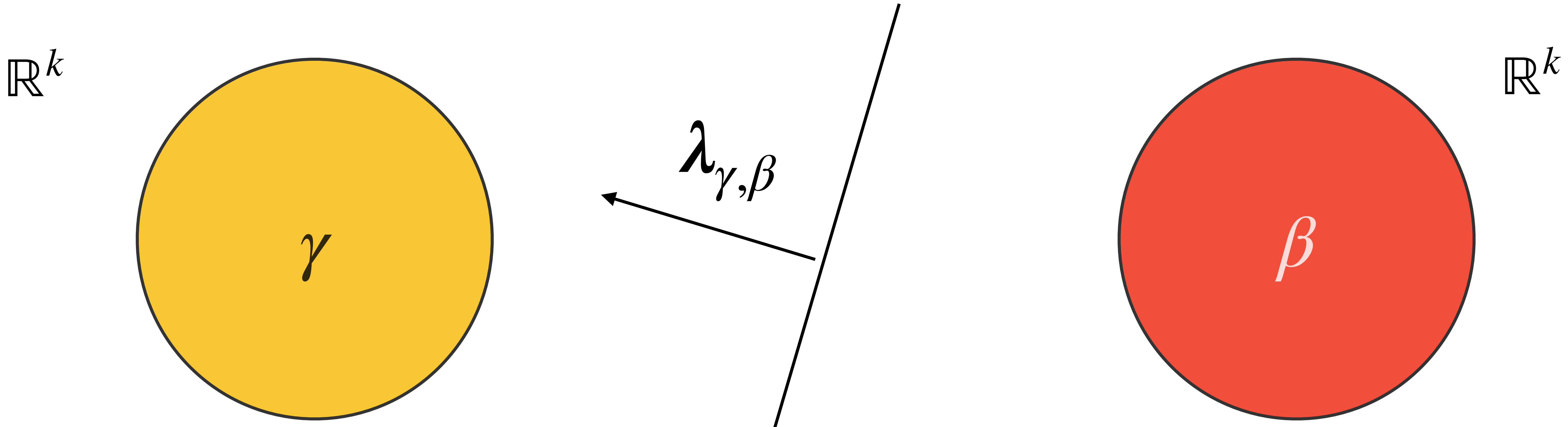
q := alphabet size

$$\text{Bias} = \|\boldsymbol{\lambda}_{\gamma,\beta}\mathbf{B}\|_{1,\infty}$$

k := arity of constraint, B_{ij} := # constraints with x_j in the i-th position

A closer look— how are the yellow and red sets defined? why does Bias work?

$$\text{Bias} = \max_{\mathbf{a}\in\{-1,1\}^n} \boldsymbol{\lambda}_{\gamma,\beta} \underbrace{\mathbf{B}\mathbf{a}^T}_{\in\mathbb{R}^{k\times 1}}$$



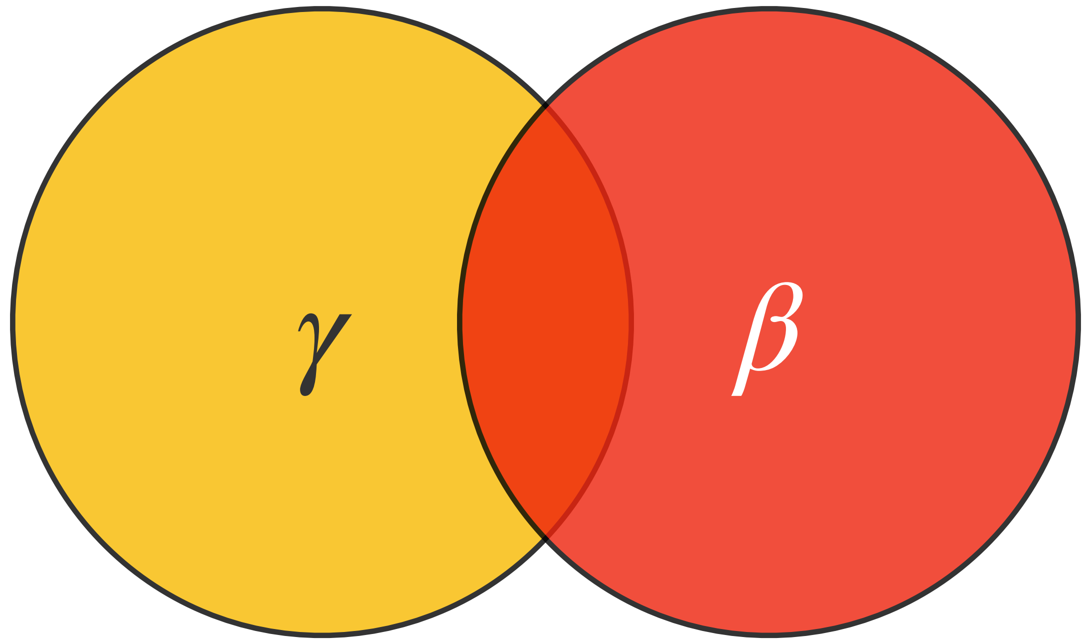
OPT $\geq \gamma \implies \exists \mathbf{a} : \mathbf{B}\mathbf{a}^T$ is in the **yellow** set

OPT $\leq \beta \implies \forall \mathbf{a} : \mathbf{B}\mathbf{a}^T$ is in the **red** set

Bias $> \tau$

Bias $< \tau$

Lower bounds



- Create indistinguishable distributions over OPT $\geq \gamma$ instances and OPT $\leq \beta$ instances

- Lower bounds follow from hardness of communication games. Unconditional!

Open questions

1. $1/2$ approximation for Max-DICUT in
   I. sublinear space
   II. random ordering
   III. multi-pass setting
2. Dichotomy theorems in sublinear space
   I. Are first order snapshots optimal in $\sqrt{n}$ space? Does the best oblivious algorithm also lead to the best $\widetilde{\Theta}(\sqrt{n})$ space algorithm?

Thank you for your attention :)