Streaming & Sketching CSPs

Madhu Sudan
Harvard University

Based on joint works with Chi-Ning Chou, Alexander Golovnev, Noah Singer, Ameya Velingker and Santhoshini Velusamy.

This Talk

- CSPs (& approximation & streaming/sketching)
- Our results
- Some Proof Ideas

Constraint Satisfaction Problems (CSPs)

- Class of infinitely many problems. $\boxed{\text{MaxCSP}(F)}$

- Specified by q, k and family $F \subseteq \{f: [q]^k \to \{0,1\}\}$.

- Instance of MaxCSP(F): $\Psi = (X_1, \ldots, X_n; C_1, \ldots, C_m)$;
  Constraint $C_j(X) = f_j(X_{i_1(j)}, \ldots, X_{i_k(j)})$; $f_j \in F$
  - $opt(\Psi) \equiv \max_{a \in [q]^n} \{\sum_j C_j(a)\} \frac{1}{m}$

- Examples:
  - MaxCut ($F = \{\oplus\}$)
  - MaxDicut ($F = \{x \wedge \bar{y}\}$)
  - Max 3SAT ($|F| = 11$), Max Exact 3SAT ($|F| = 4$), Max q-Colorability.

Constraint Satisfaction Problems (CSPs)

- "Special Case": Boolean MaxCSP(F):
  - $q = 2$;
  - Constraints $C = f(X_1, \overline{X_2}, \dots, \overline{X_k})$
    - Constraints applied to literals.
  - Warning: MaxCSP$(F) \neq$ Boolean MaxCSP(F)
  - $\forall F\ \exists G$ s.t. Boolean MaxCSP$(F) =$ MaxCSP(G). (converse not true. MaxCut, MaxDicut …)

$\sigma_1\ \sigma_2 \ldots \quad \tau_1 \ldots \tau_2$

$S(\sigma)$ $\quad S(\tau)$ $\quad S' = f(S, T)$

Streaming & Approximation

- Streaming input
  - Streaming: $s(n)$-space algorithm – gets one constraint at a time.
  - Sketching: Maintains sketch $S(\sigma)$ with $|S(\sigma)| \leq s(n)$;
    - Restriction: $\big(S(\sigma), S(\tau)\big) \rightarrow S(\sigma \circ \tau)$
  - Space milestones: polylog, sqrt, or (nearly-)linear.
- Approximations:
  - Usual notion: α-approximation:
    - Output v s.t. $\alpha \cdot opt(\Psi) \leq v \leq opt(\Psi)$
  - Refined notion: (γ, β)-distinguishability:
    - Output Yes if $opt(\Psi) \geq \gamma$, No if $opt(\Psi) \leq \beta$
  - Equivalence: $\alpha = \min_{\gamma} \max_{\beta} \beta/\gamma$

Trivial Approximations:

- $\tilde{O}(n)$-space, $1 - o(1)$-approx.
  - Maintain $\tilde{O}(n)$-constraints. Solve optimally on those using exponential time.
- $O(1)$-space, $\rho_{\min}$-approx.
  - Defn: $\rho_{\min}(F) := \min_{\Psi} \{\mathrm{val}_\Psi\}$
  - Notes: Usually $\rho_{\min} > 0$ (unless $0 \in F$)
  - For Boolean CSPs (on literals)
    - $\rho_{\min}(f) = 2^{-k} \cdot |f^{-1}(1)|$ (= value of random assgmt).
    - E.g. $\rho_{\min}(\mathrm{MaxCut}) = \frac{1}{2}$
- Key question: (When) can we do better than trivial?

Why study CSPs

- Contain some problems of direct interest
  - Max Cut, Max Dicut, Max colorability
- Allow possibility of classification!
- Highlight general algorithms
  - Norm approximations (already used)
  - Local Exploration
  - Crude snapshots
- Identify Phenomena:
  - No $2^{\sqrt{\log n}}$-space algorithms?

Brief History

- 2011 Bertinoro W'shop (P. 45): "We know nothing +/-"
- 2015-19: Lower bounds for MaxCut [KKS,KKSV,KK]
  - Kapralov-Krachun: $\frac{1}{2}+\epsilon$-approximation requires $\Omega(n)$-space (in n vertex graph) (streaming).
- 2017-20: Algorithms for Max DiCut, Max SAT
  - Guruswami-Velingker-Velusamy: DiCut
  - Chou-Golovnev-Velusamy: DiCut, Max SAT
- 2020: Sketching Classification
  - Chou-Golovnev-Velusamy: Classify all Boolean MaxCSP with $k = q = 2$

Our Results

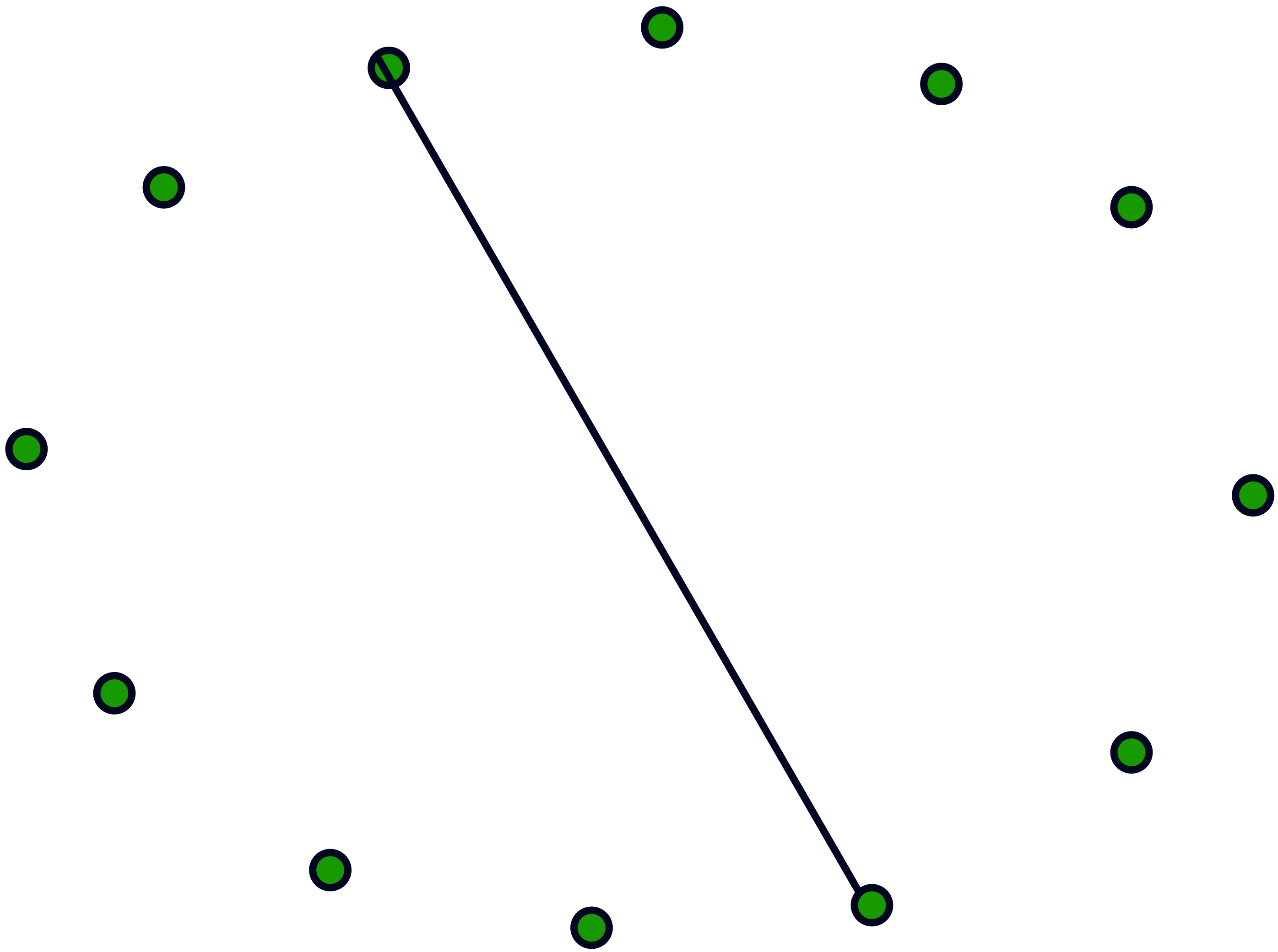
- [CGSV21]:
  - Dichotomy for sketching (polylog vs. sqrt)
  - Polylog space algorithms for infinitely many CSPs
  - $\Omega(\sqrt{n})$ space lower bounds for broad classes ("one-wise ind.", "padded one-wise ind.").
- [CGSVV 21]: Linear space lower bounds for subclass of "one-wise-ind".
  - Pins approximability of all MaxCSP(F) to within q-factor (trivial alg vs. linear space).
- [SSV21]: No sublinear algorithms for "Ordering CSPs"

Proof Ideas

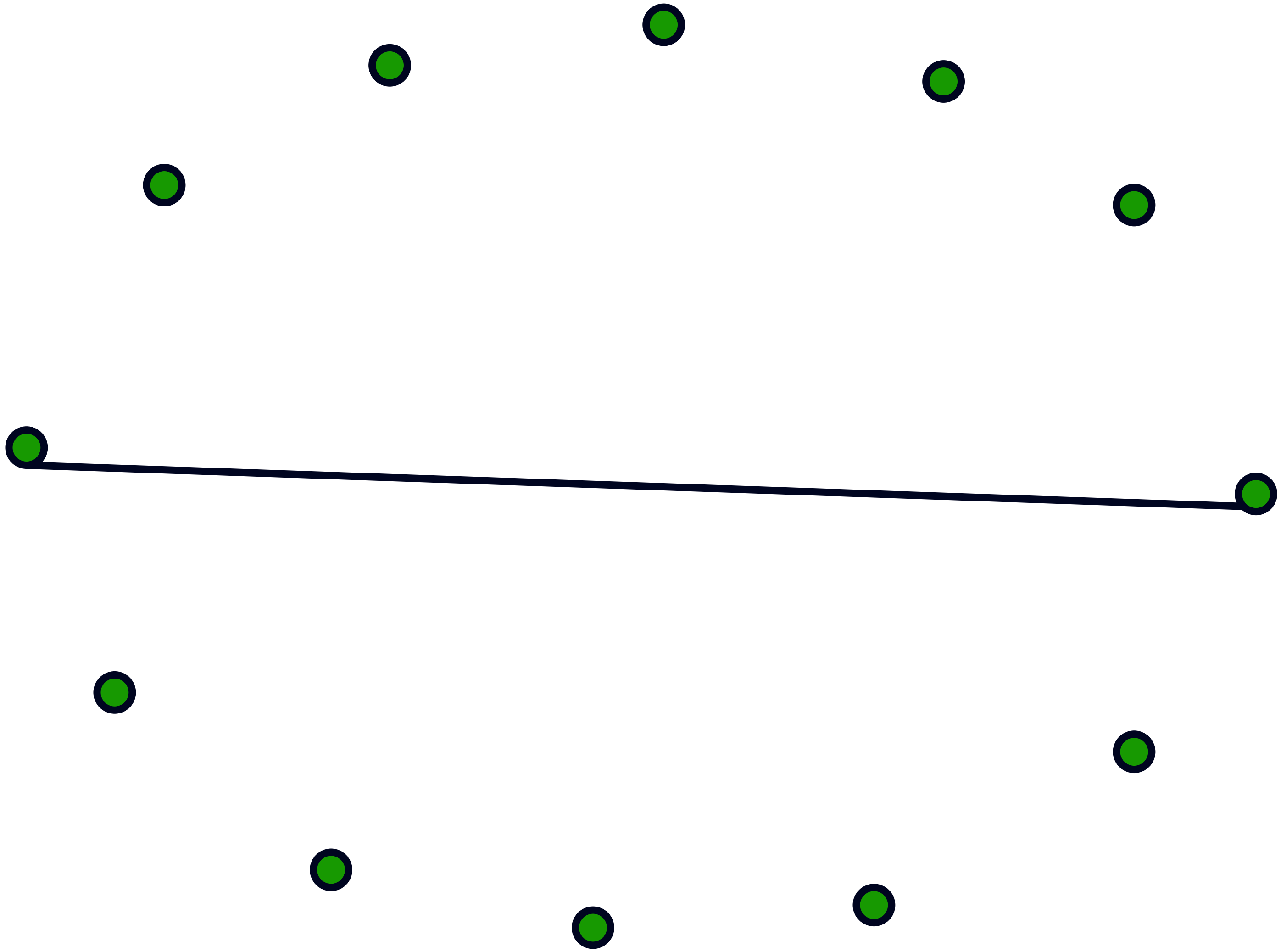
Max Cut Lower Bound

- Hard distributions:
  - YES: Random union of matchings crossing hidden bipartition
  - NO: Random union of matchings.
- Analysis:
  - Divide long stream into O(1) smaller substreams – each substream=matching.
  - Algorithm learns nothing in any single substream $\Leftarrow$ Boolean Hidden Matching Lower Bound [GKKRW]
  - Hybrid argument to combine O(1) substreams.
  - Yields $\Omega(\sqrt{n})$ space lower bound to beat trivial approximation.
- $\Omega(n)$ lower bound more complex – omitted.

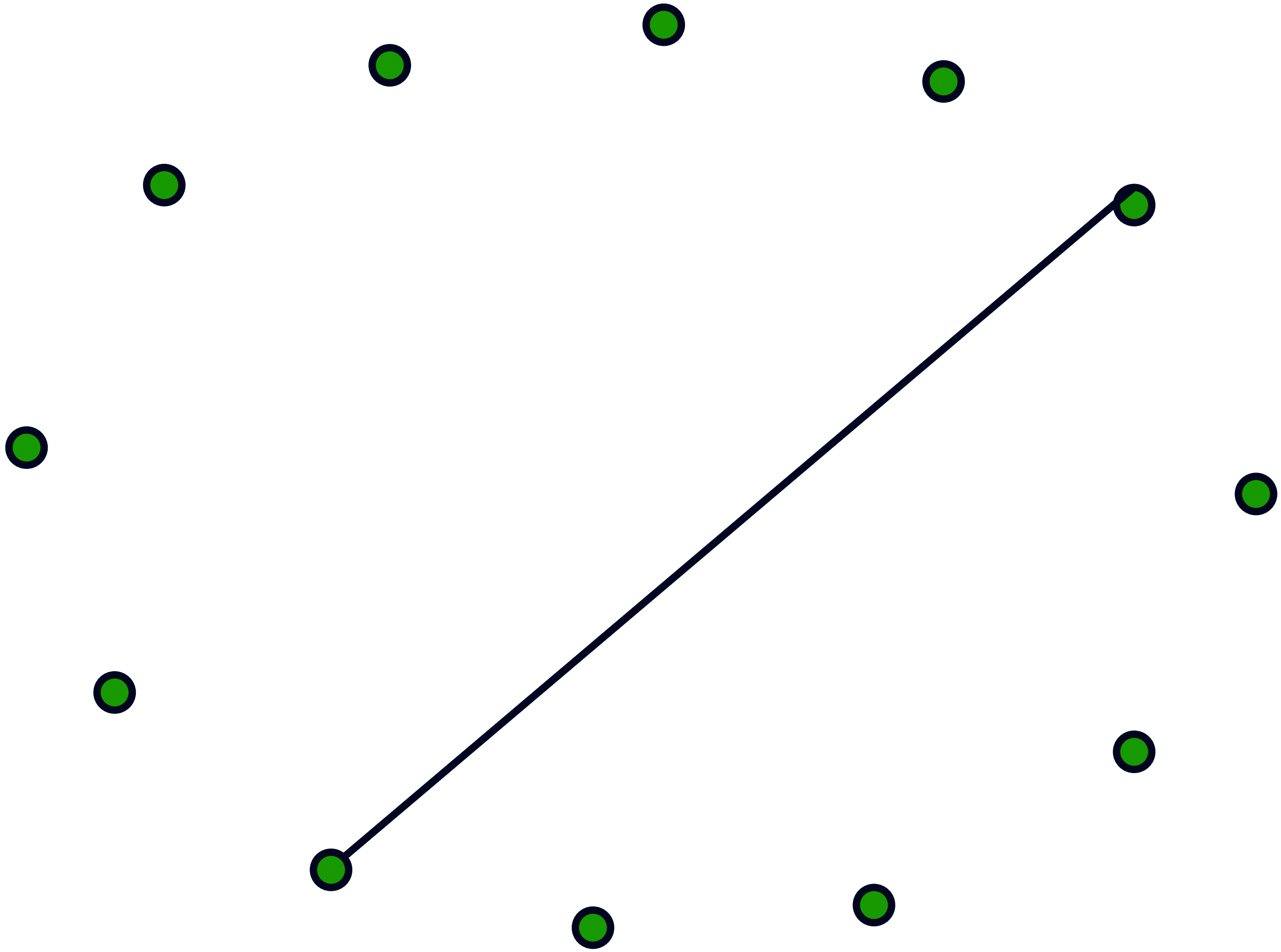
Streaming MaxCut



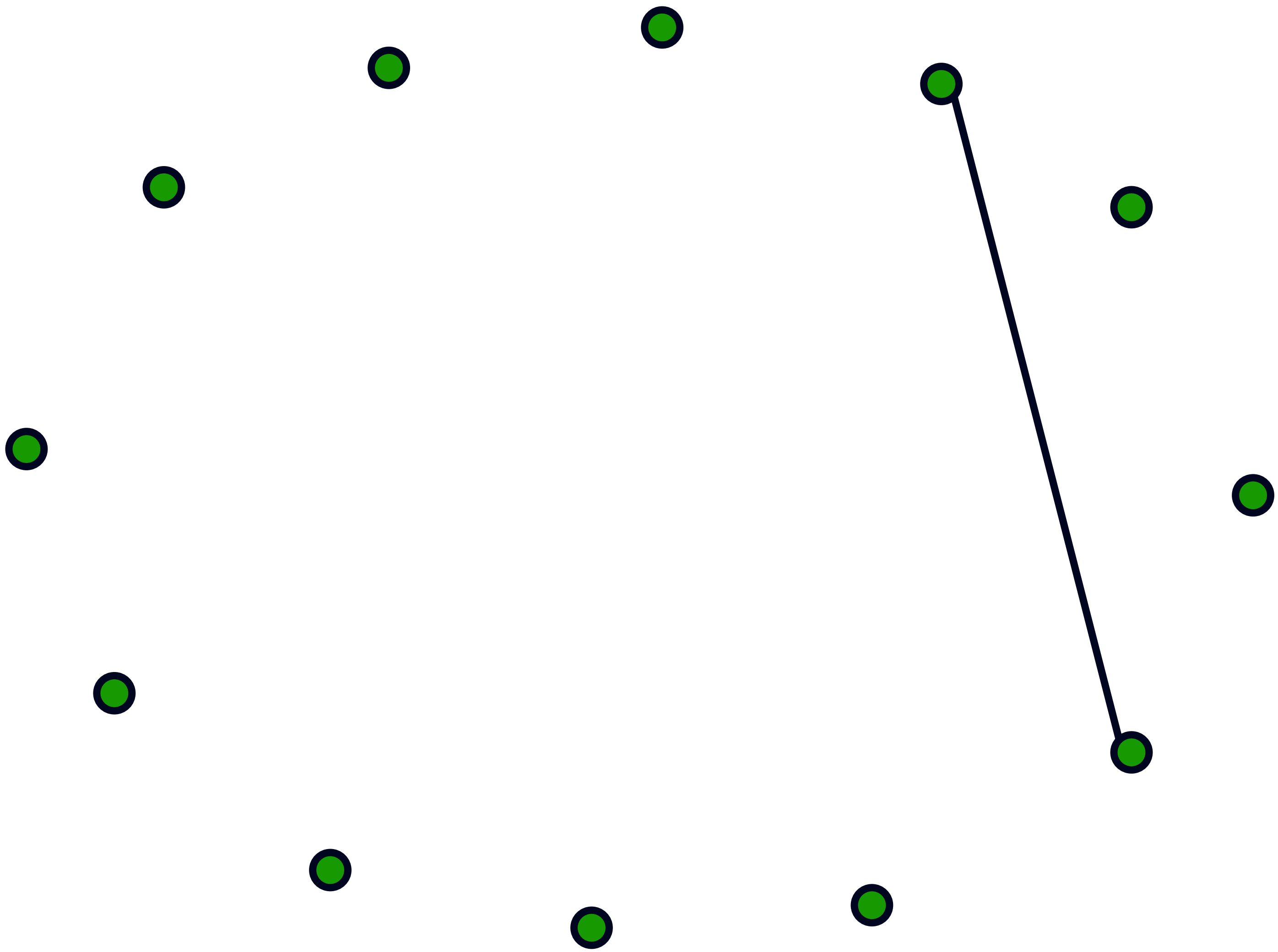
Streaming MaxCut



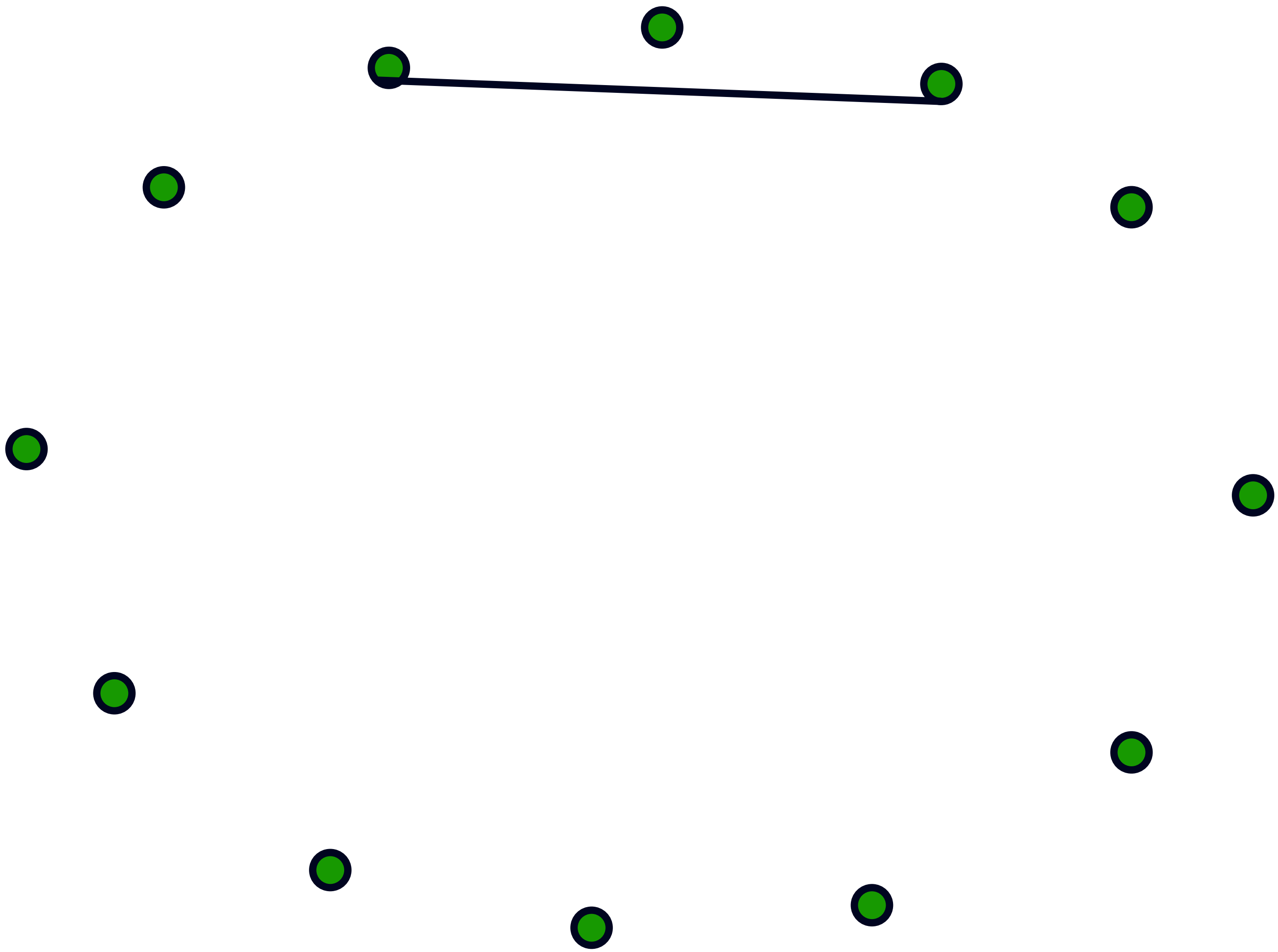
Streaming MaxCut



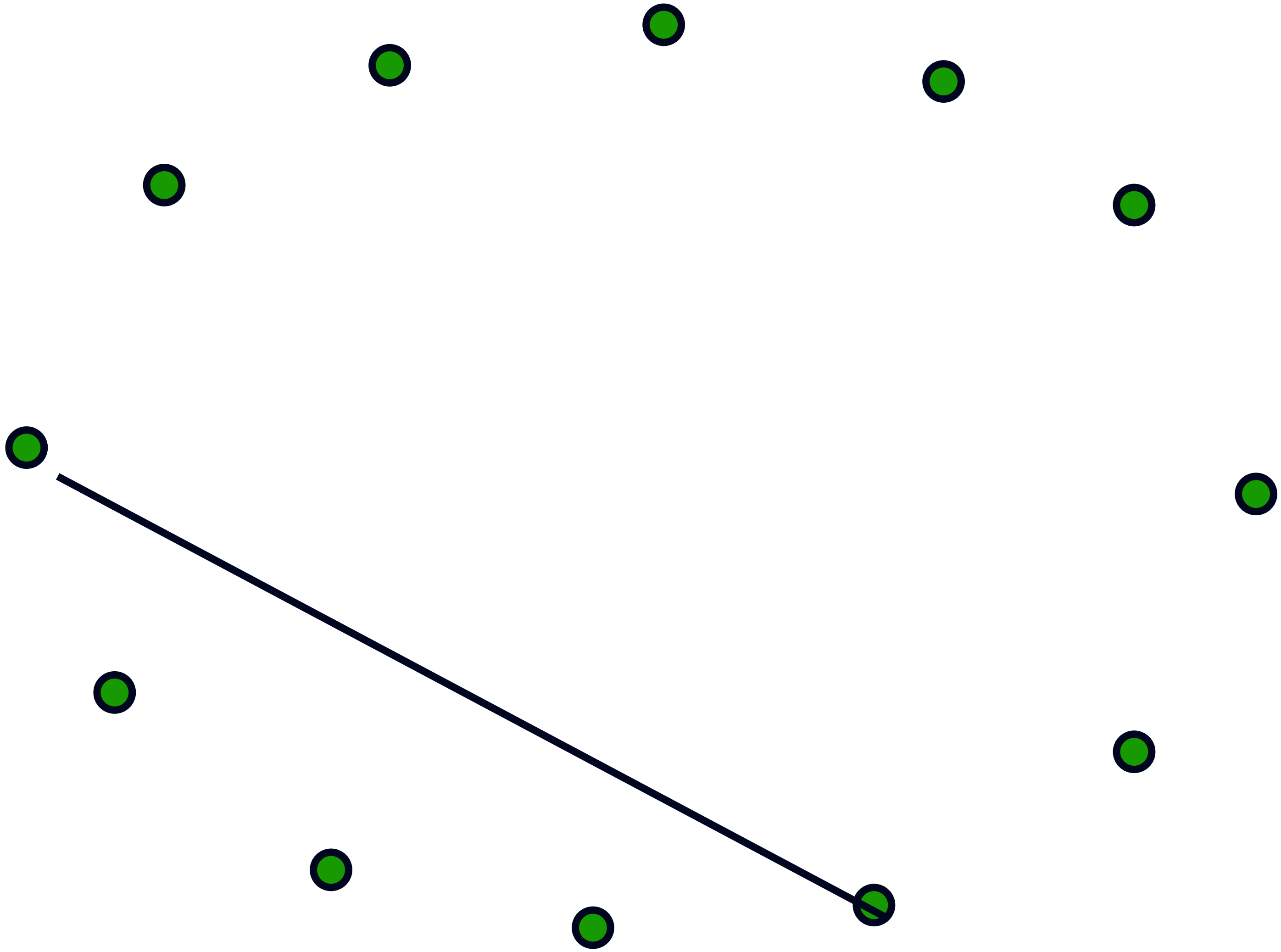
Streaming MaxCut



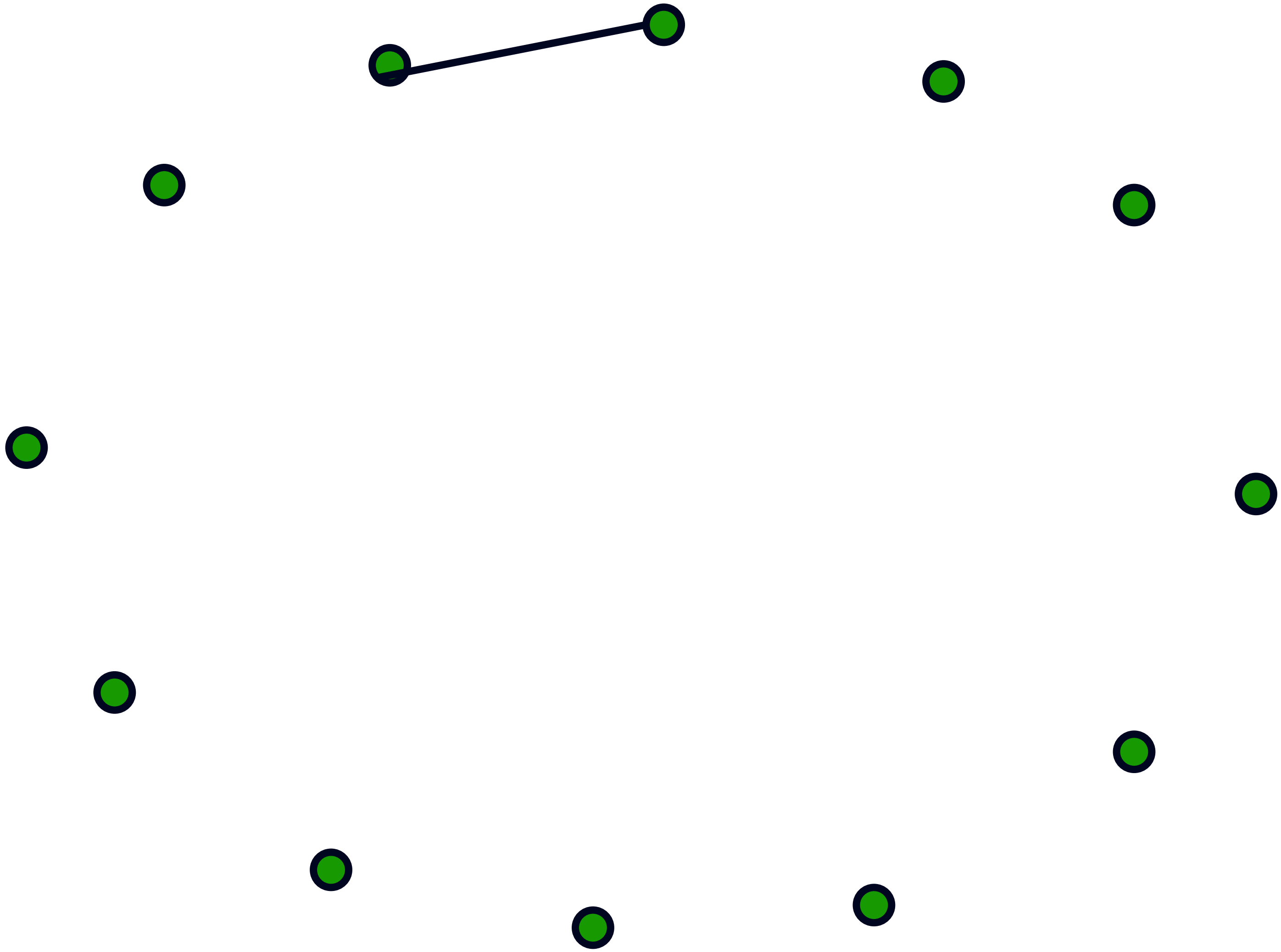
Streaming MaxCut



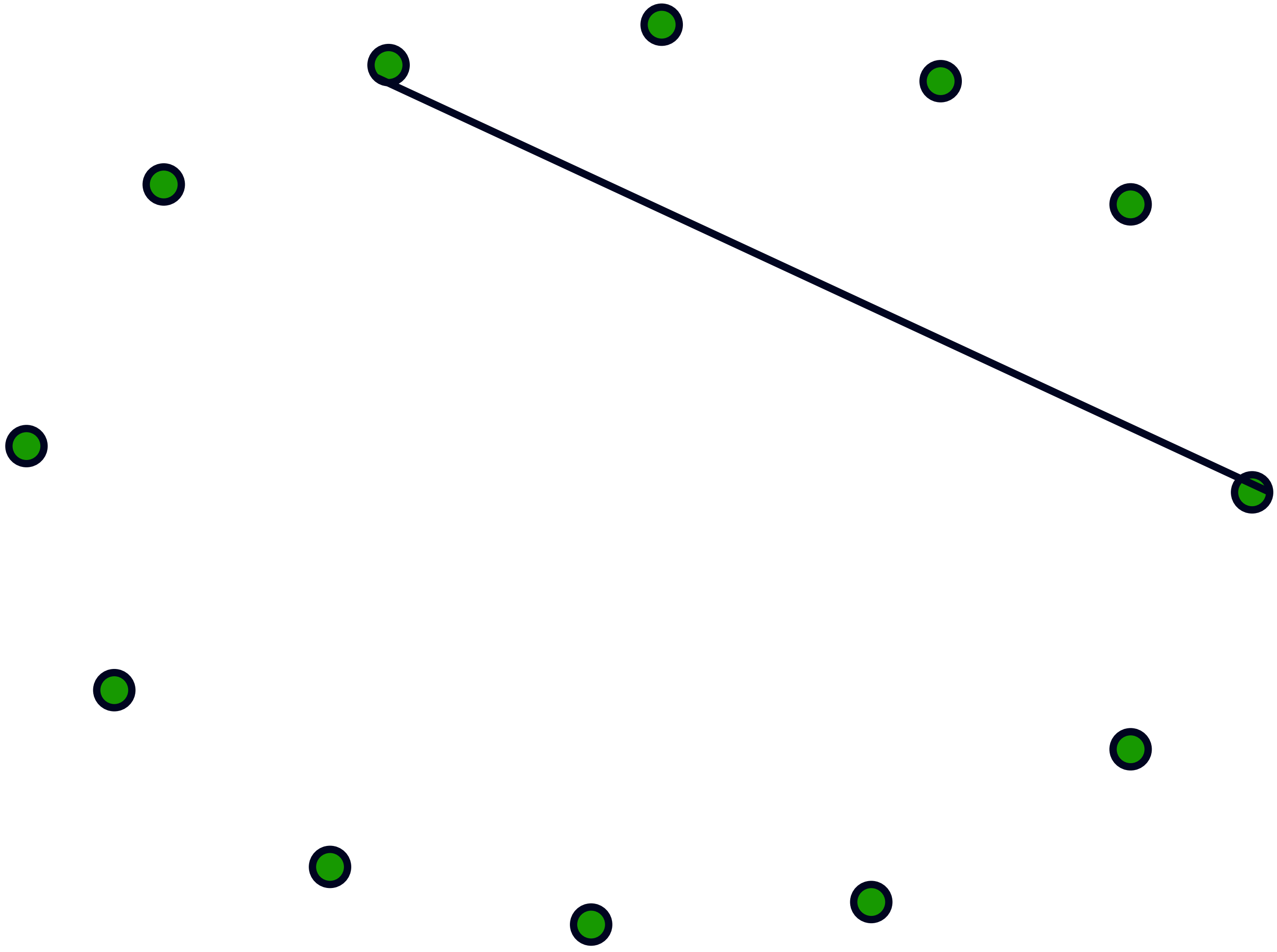
Streaming MaxCut



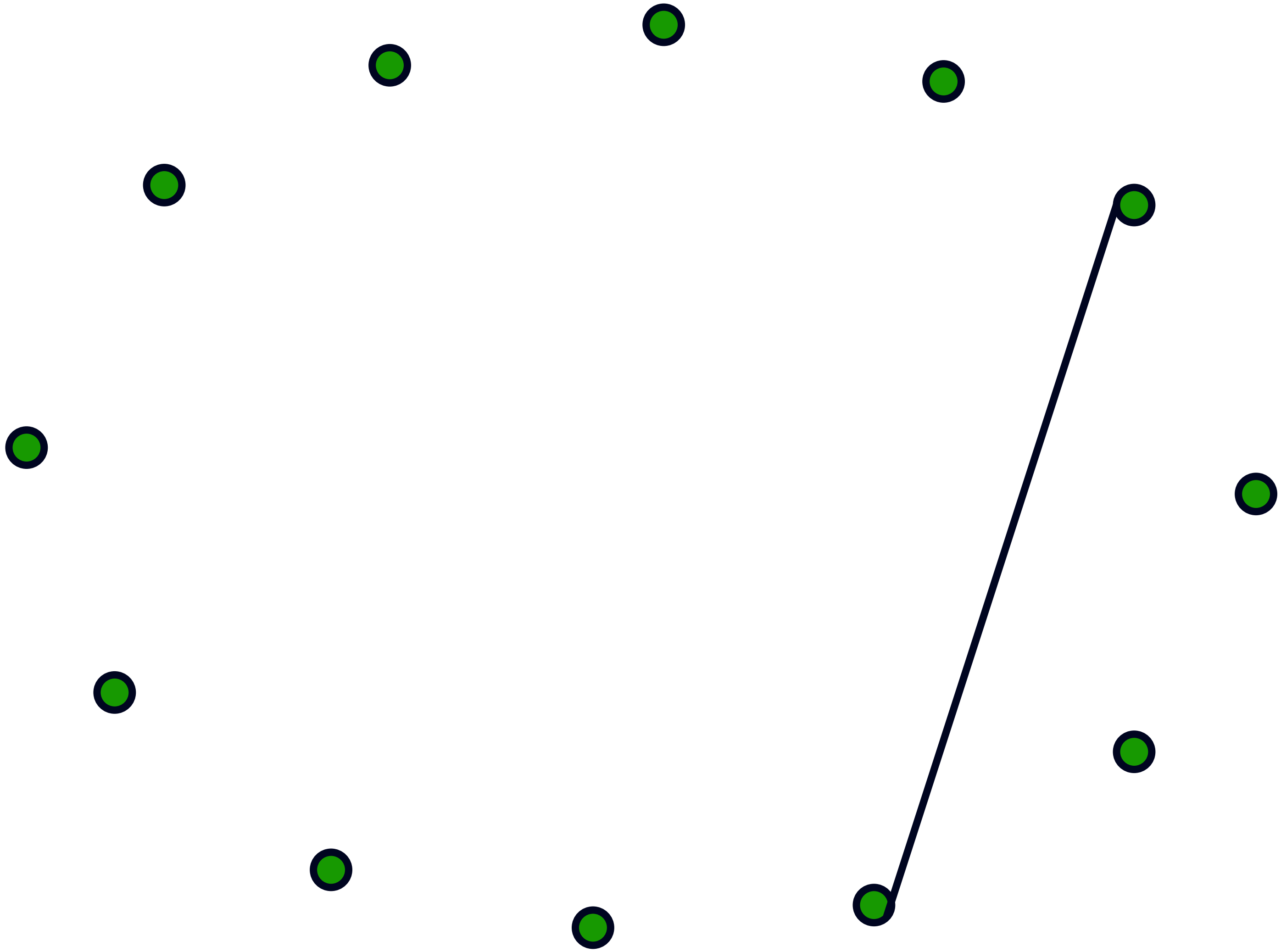
Streaming MaxCut



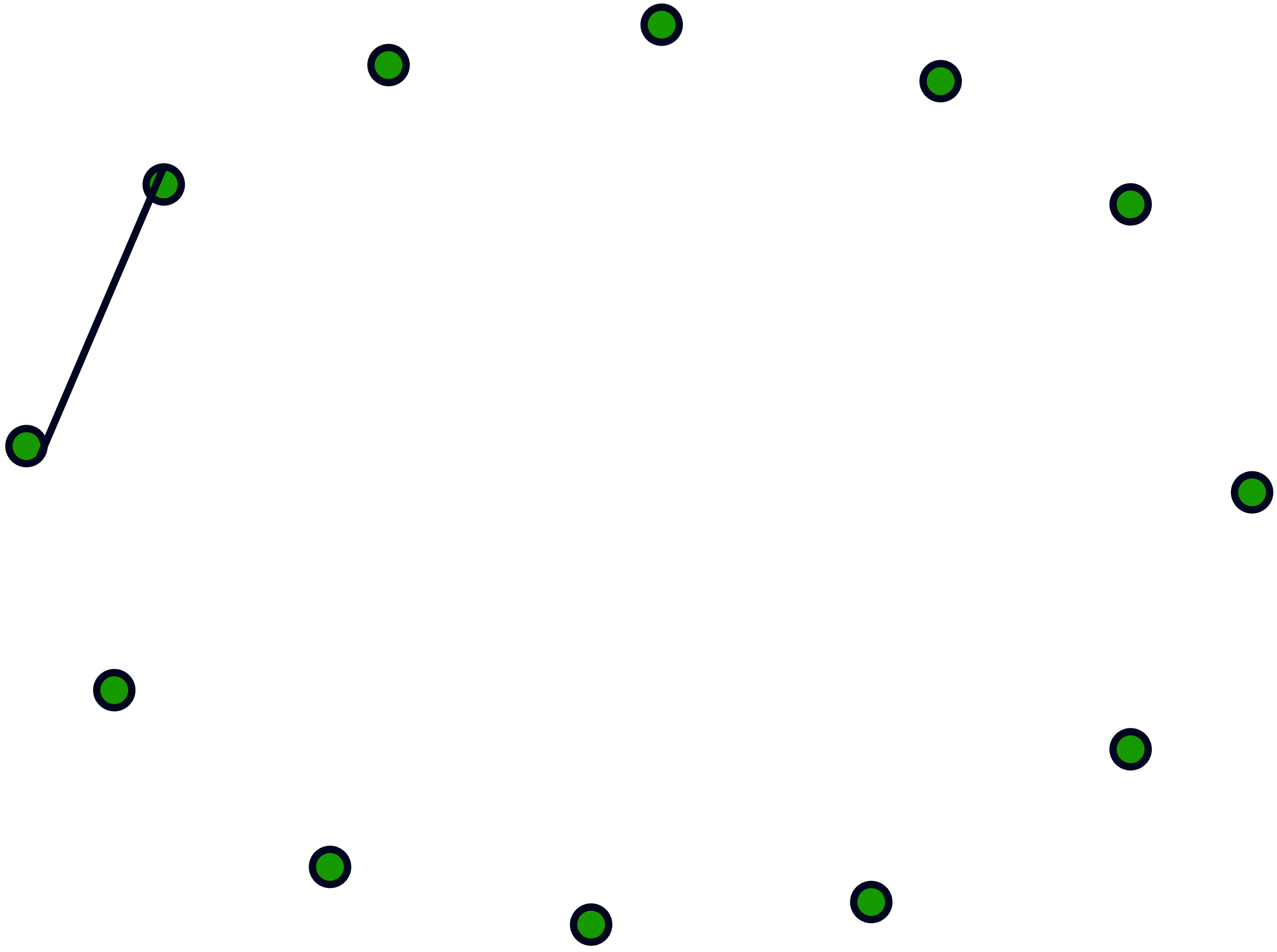
Streaming MaxCut



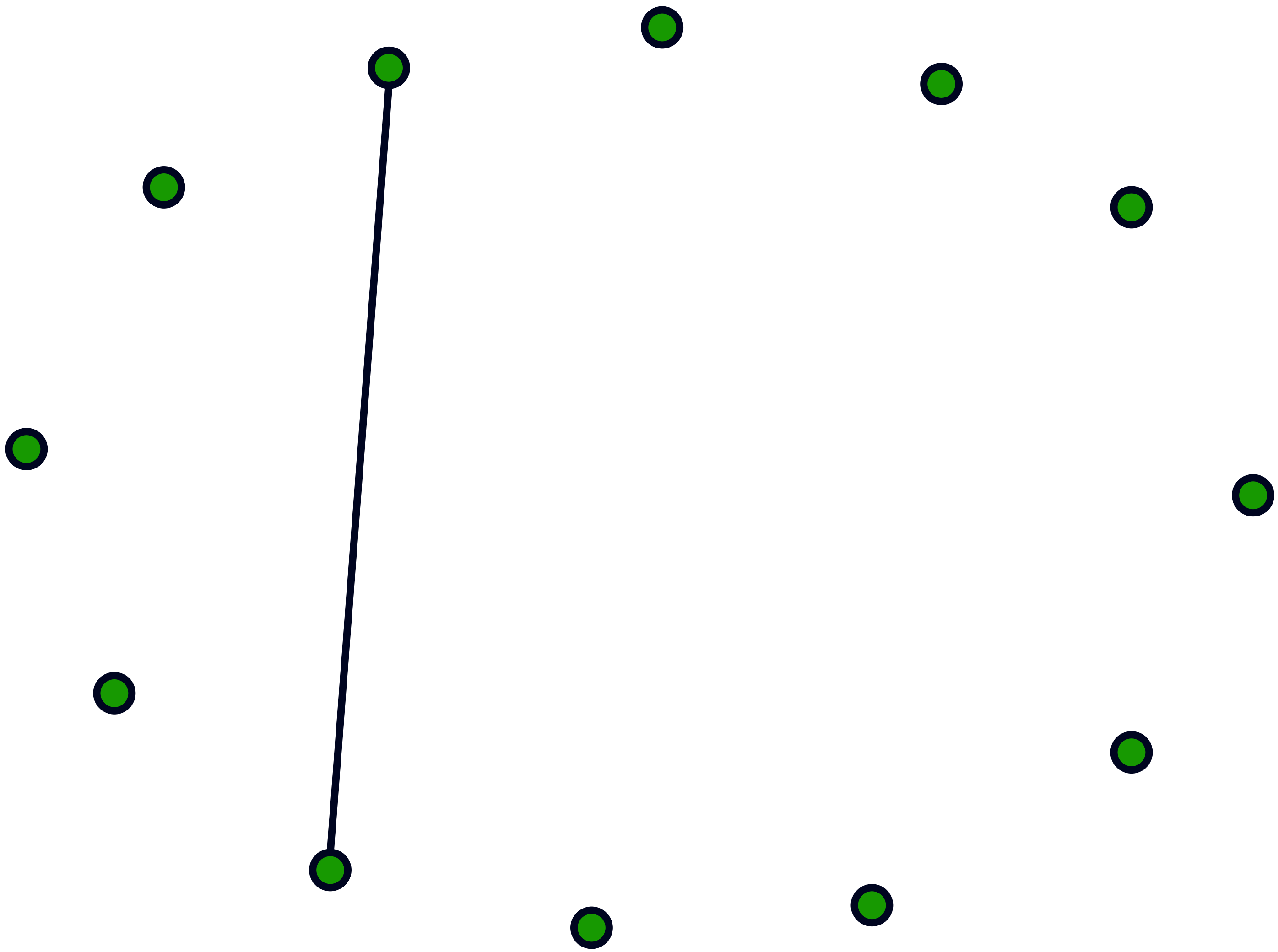
Streaming MaxCut



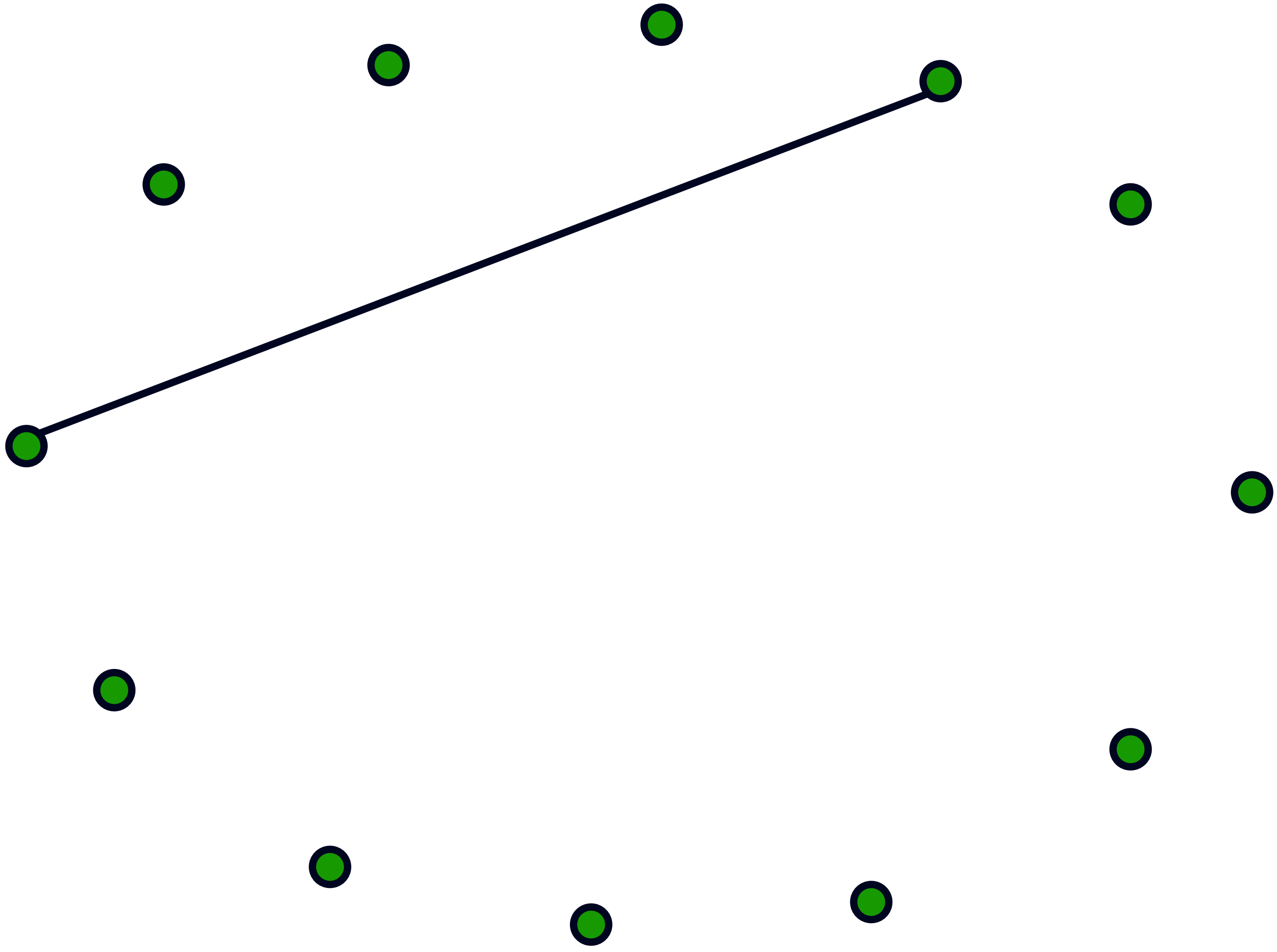
Streaming MaxCut



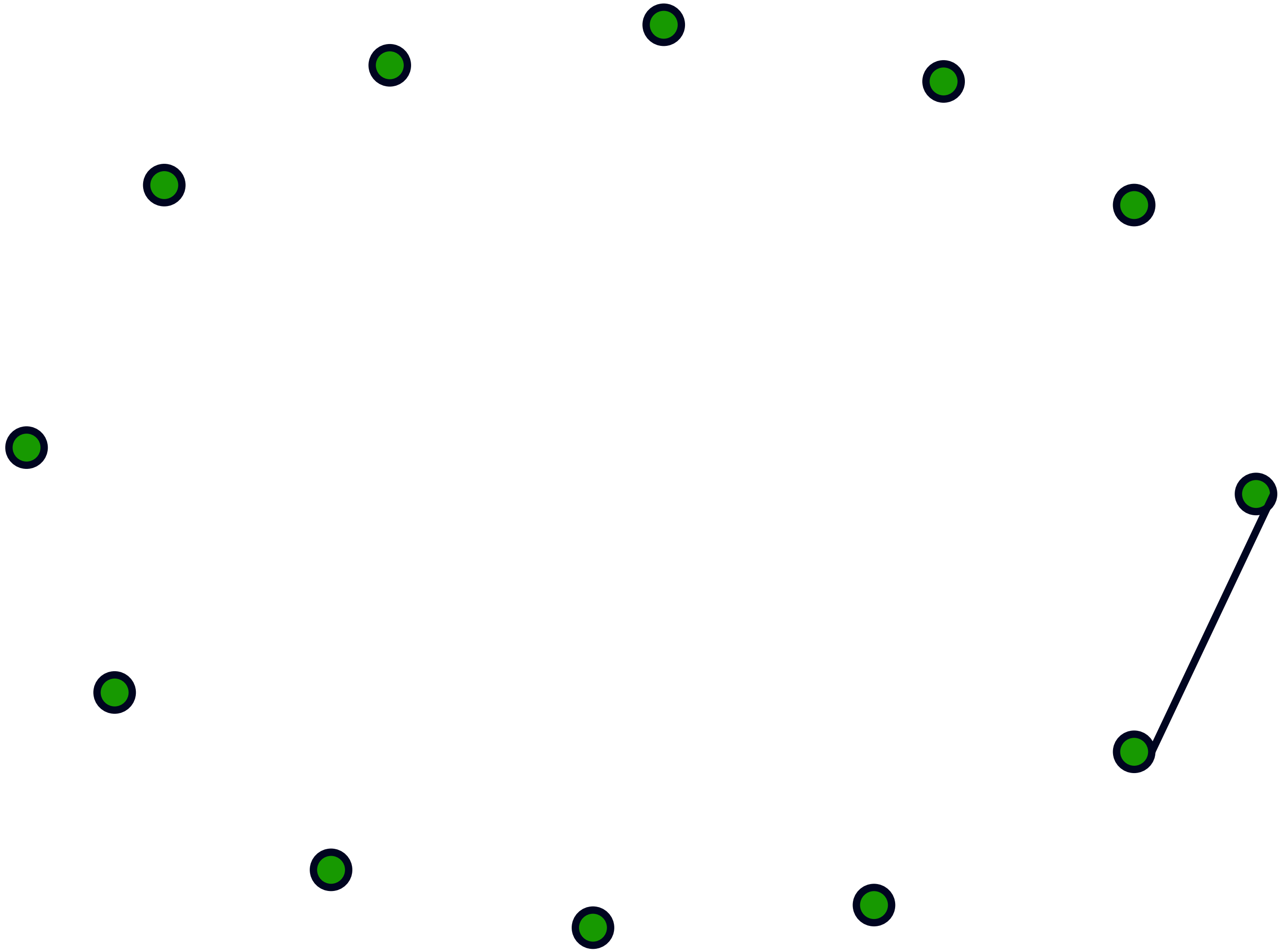
Streaming MaxCut



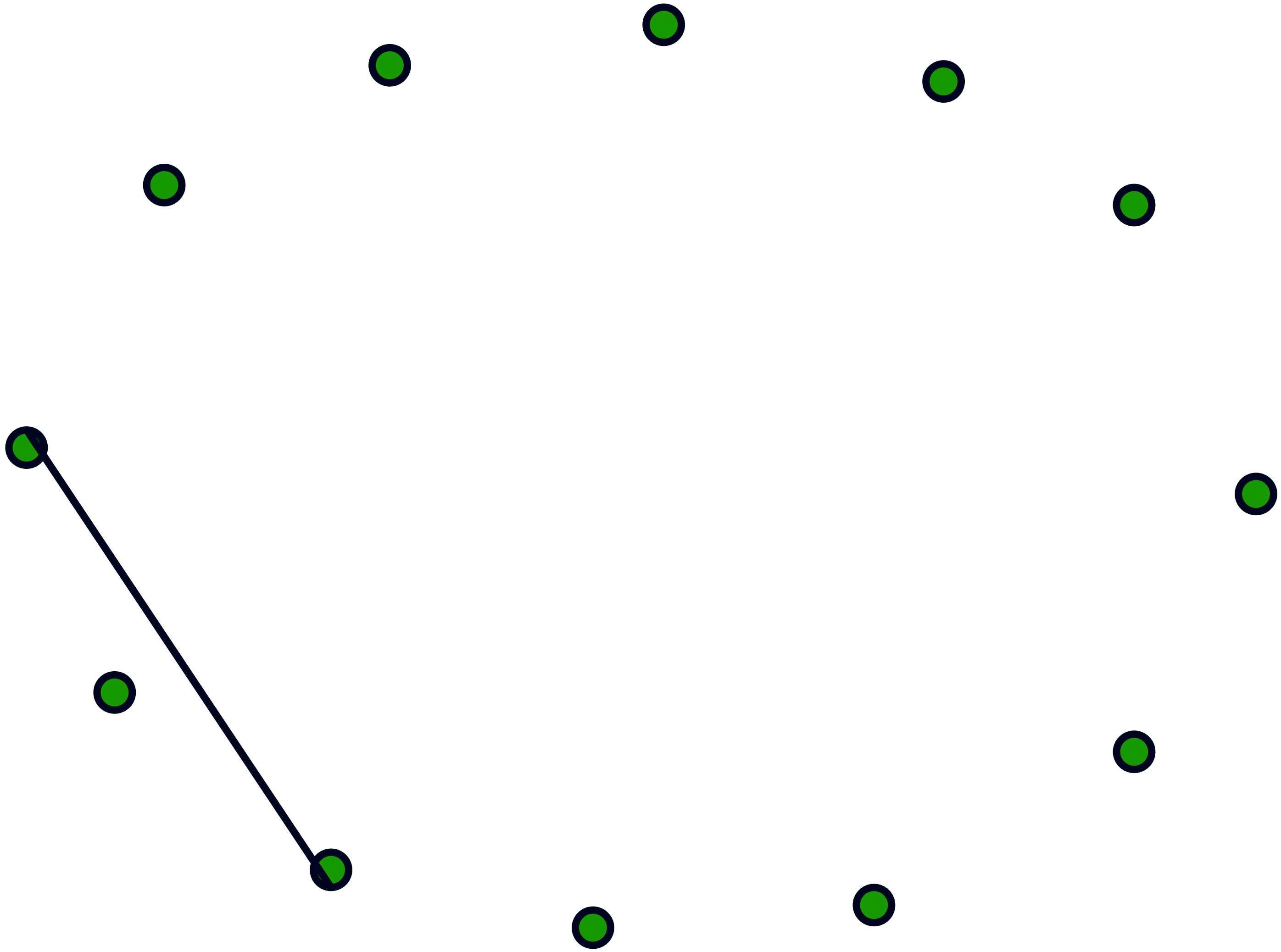
Streaming MaxCut



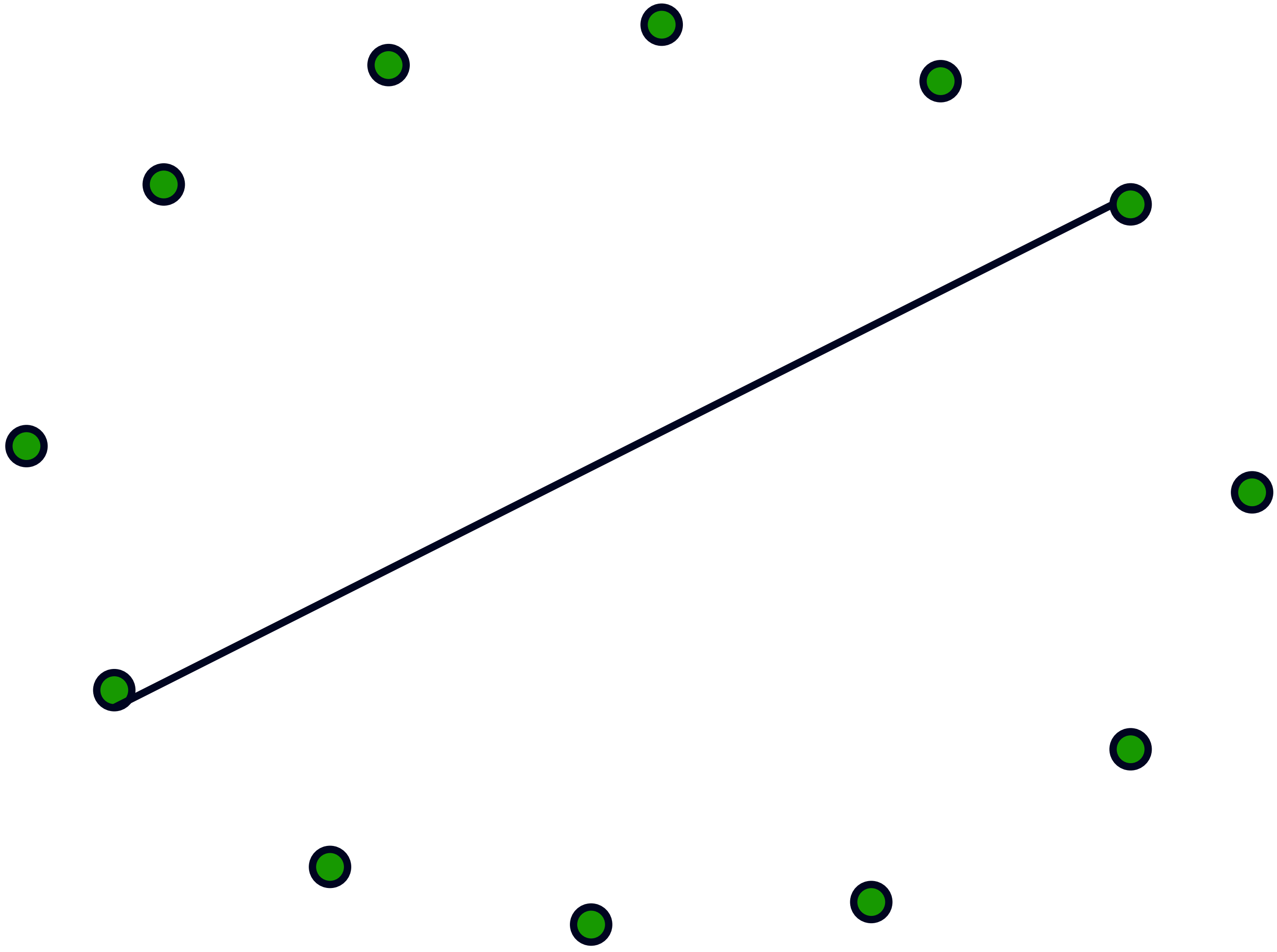
Streaming MaxCut



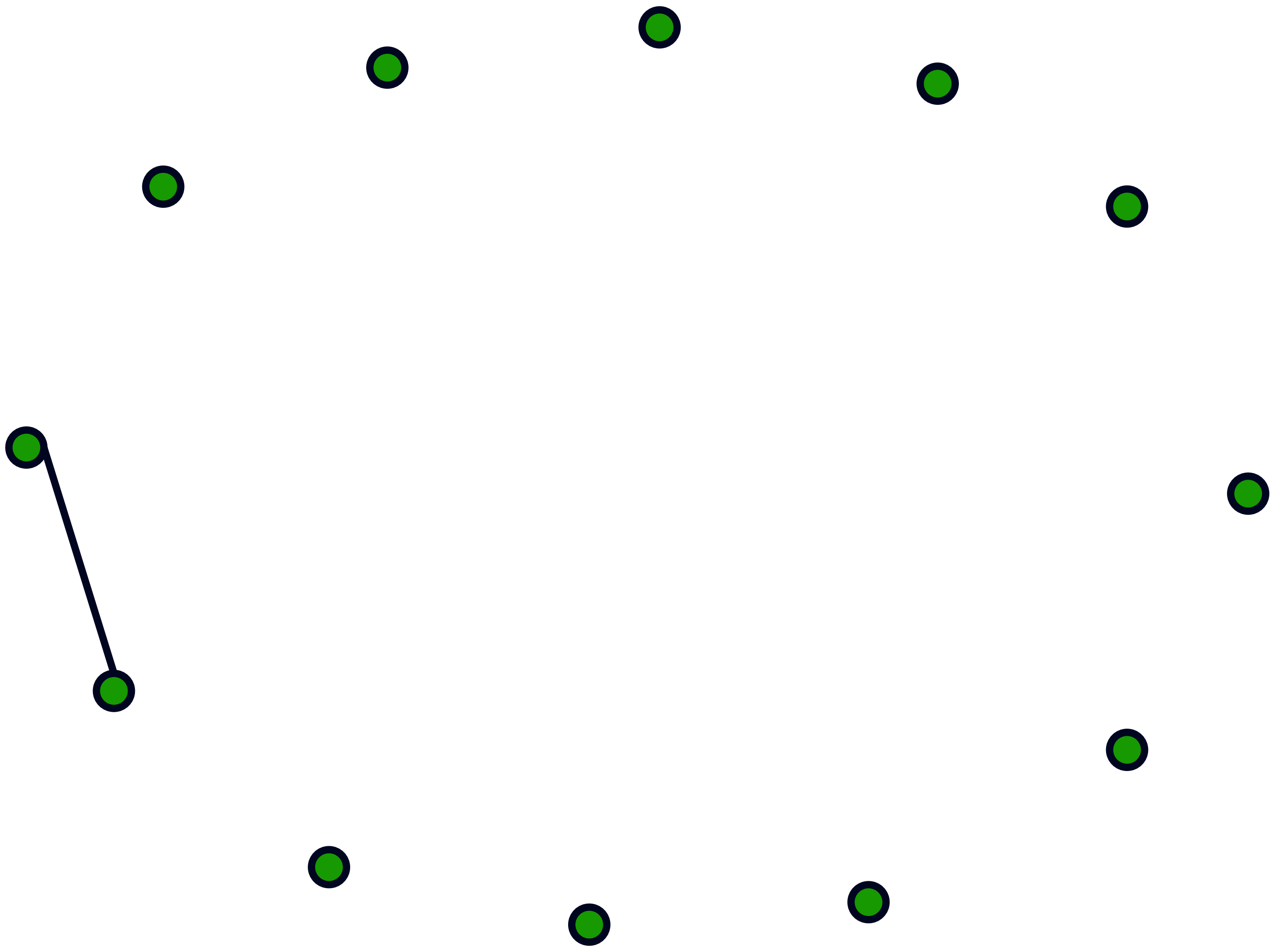
Streaming MaxCut



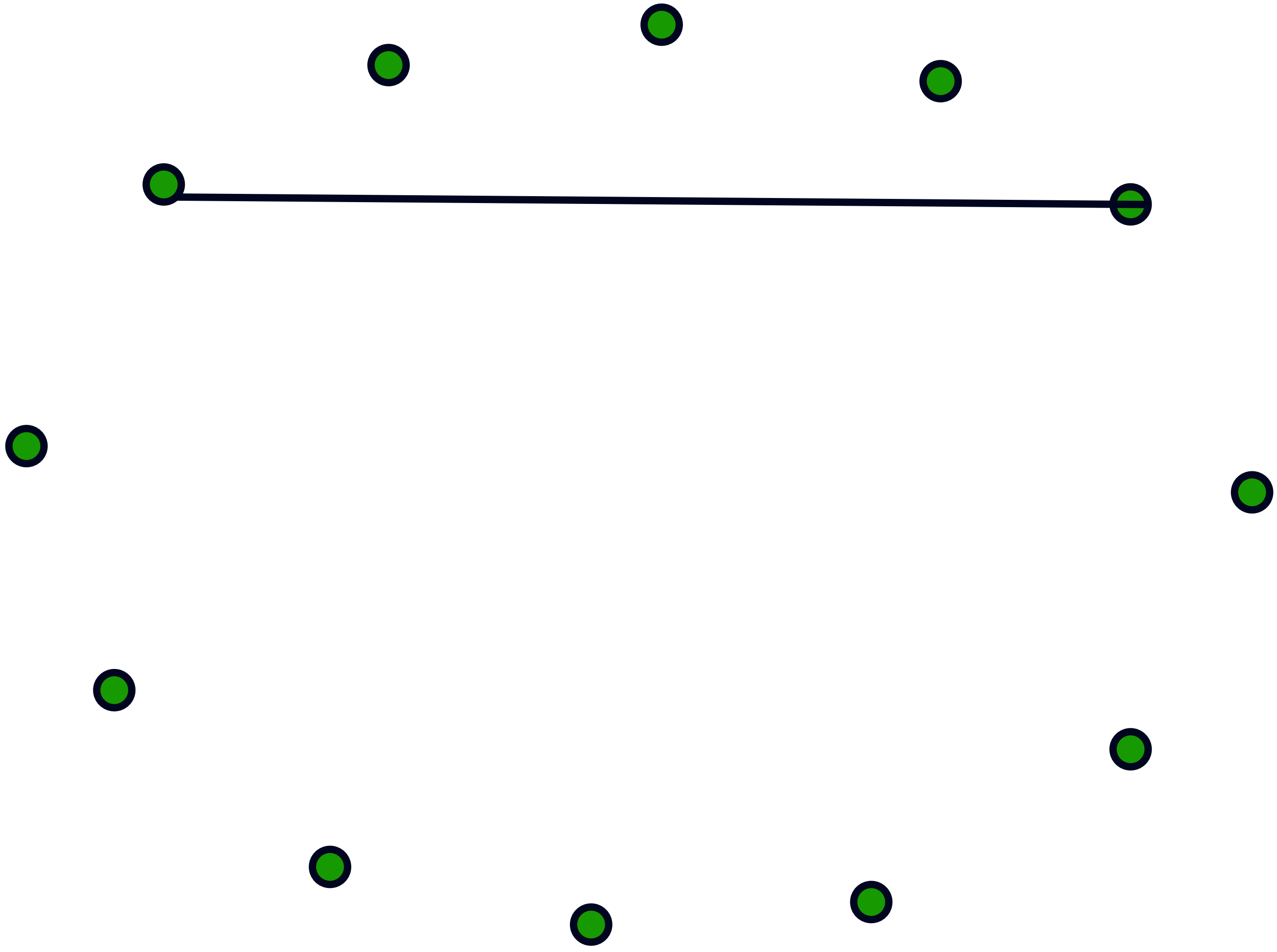
Streaming MaxCut



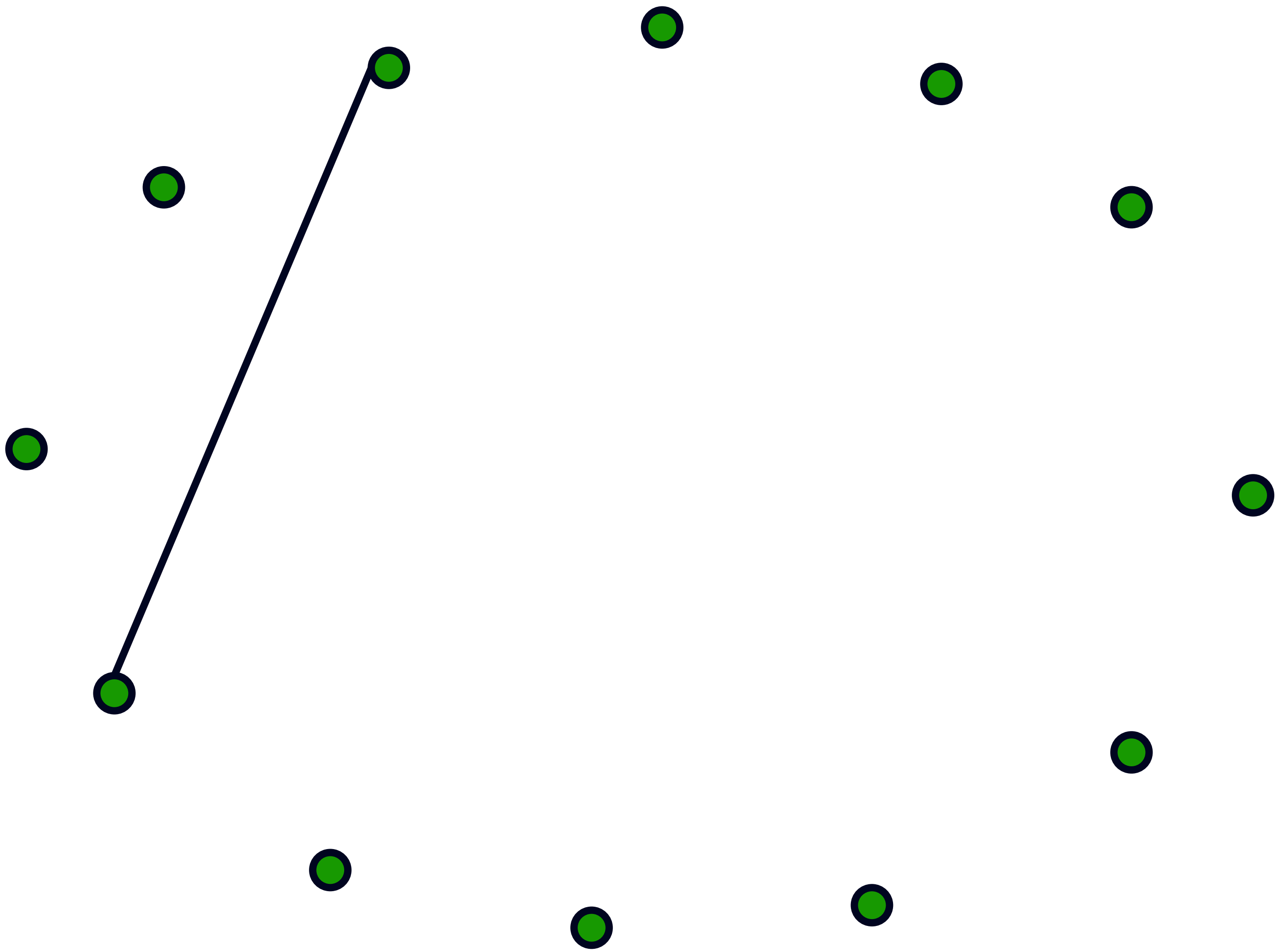
Streaming MaxCut



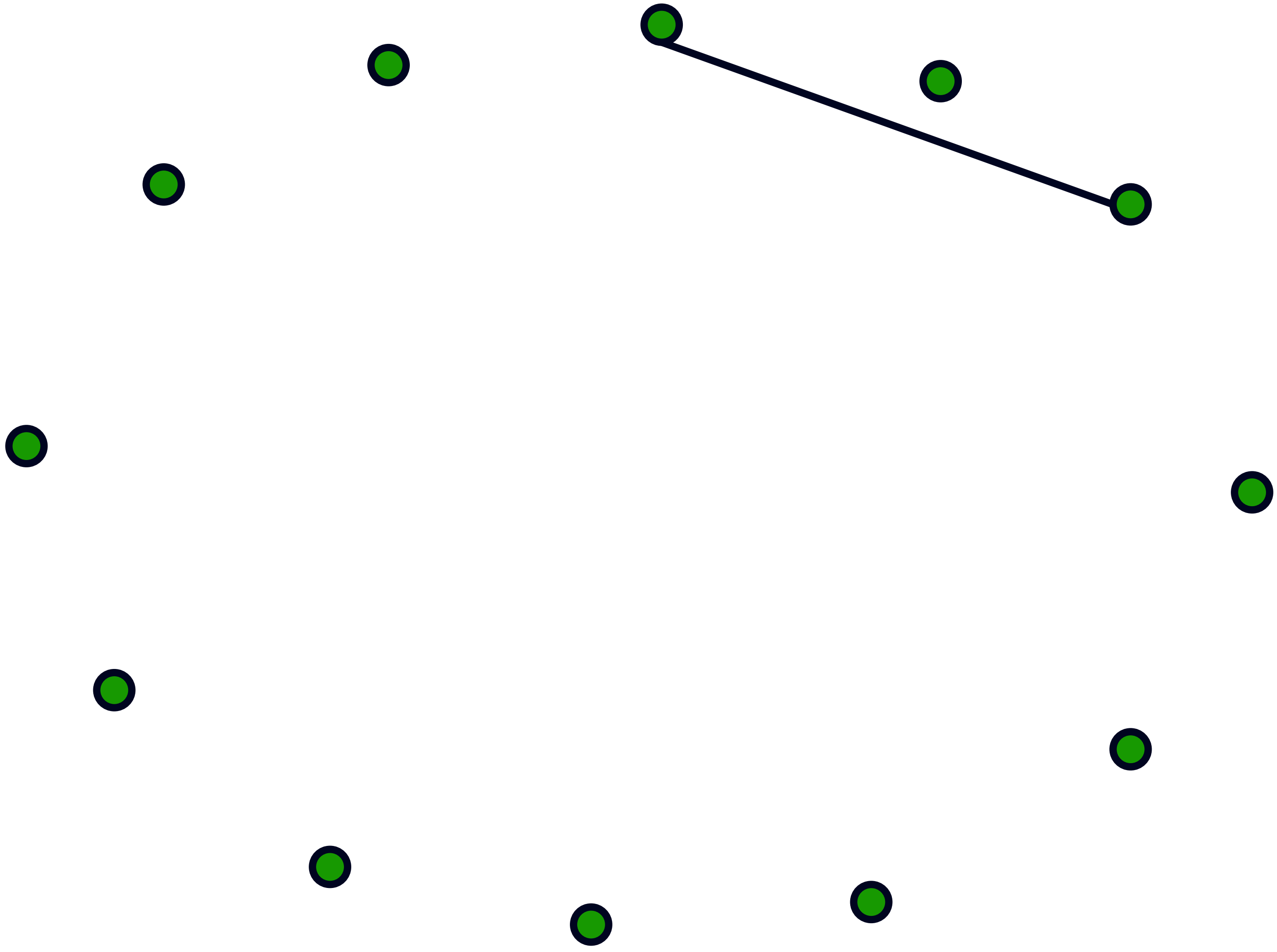
Streaming MaxCut



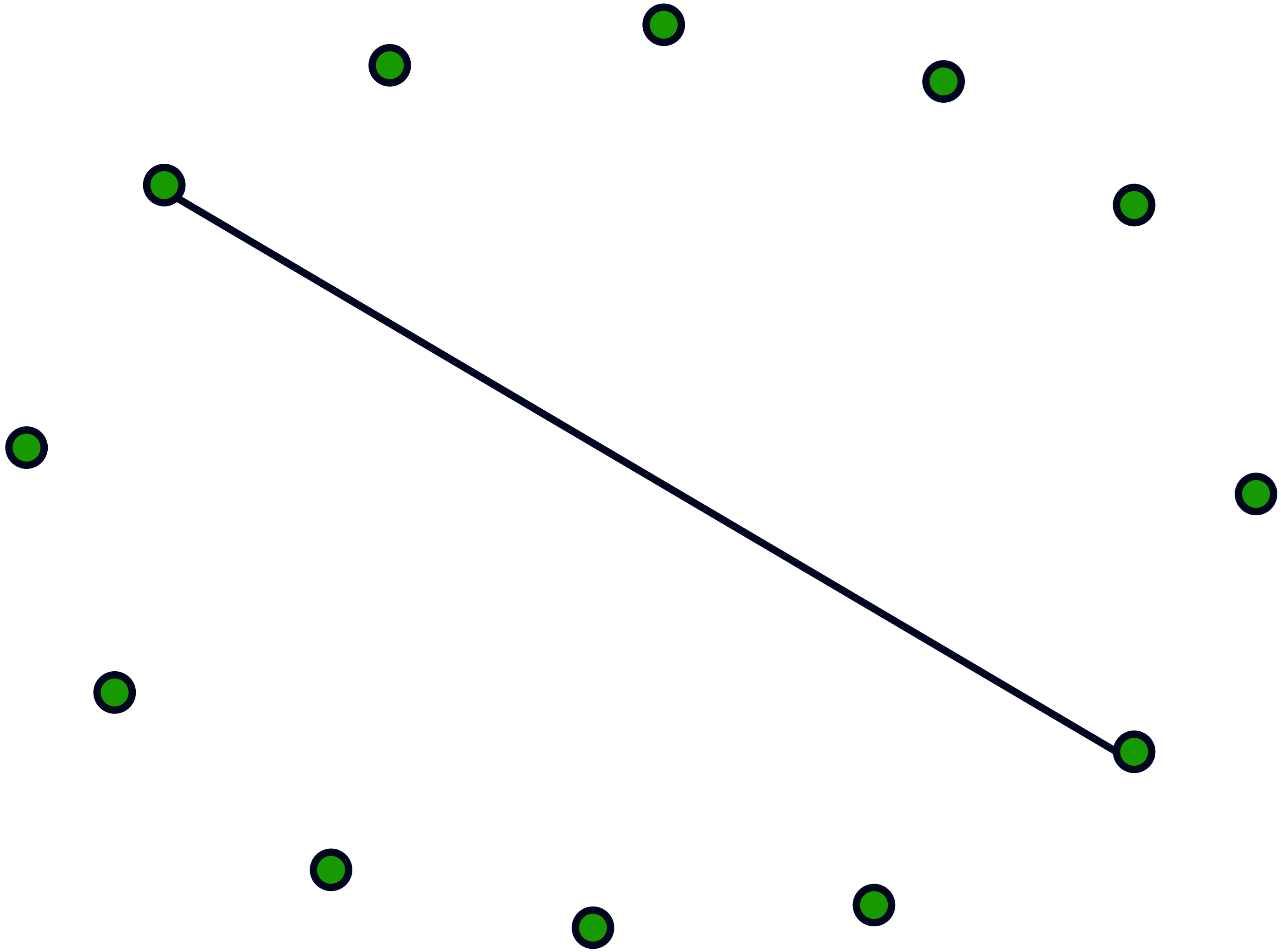
Streaming MaxCut



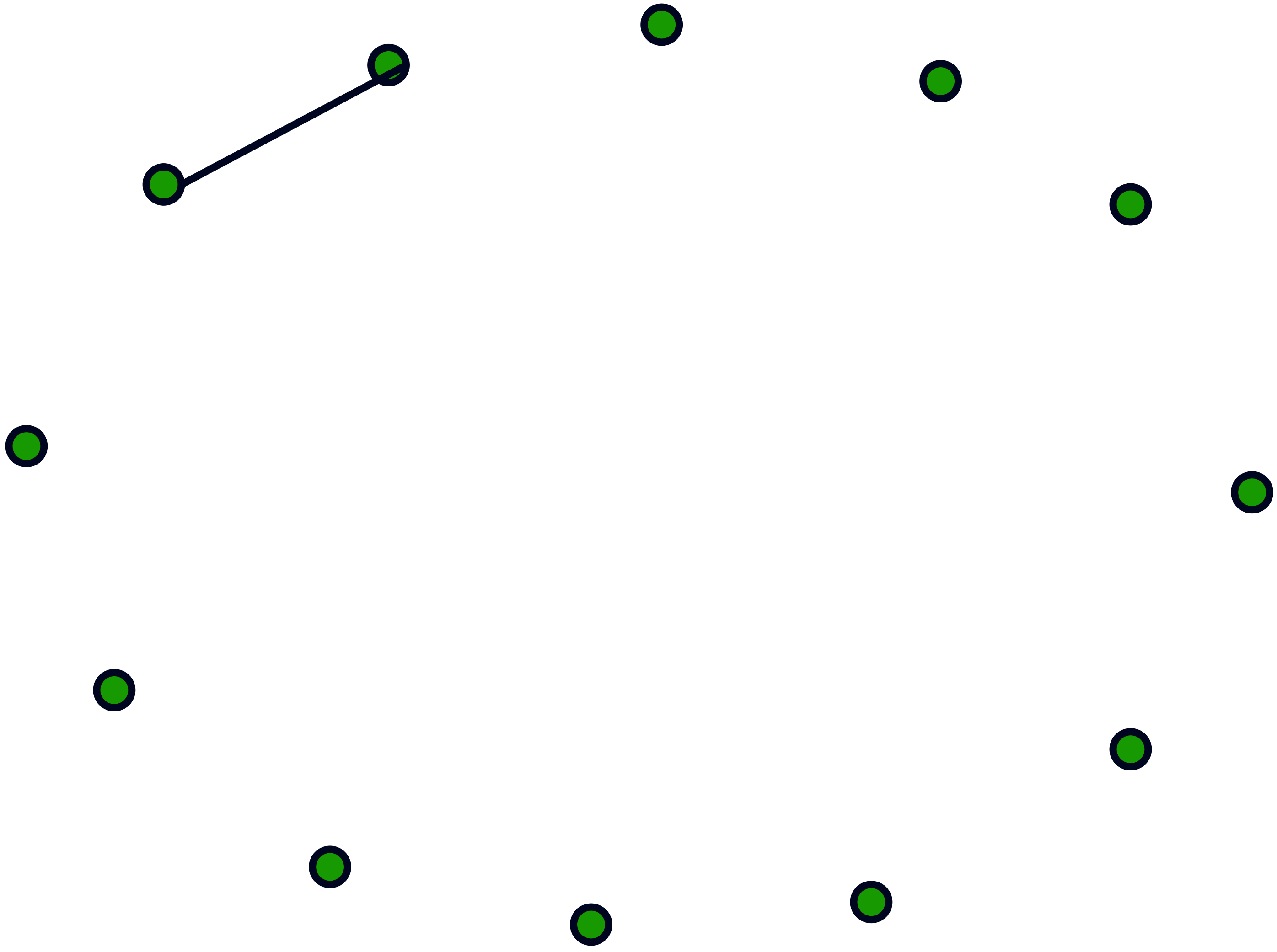
Streaming MaxCut



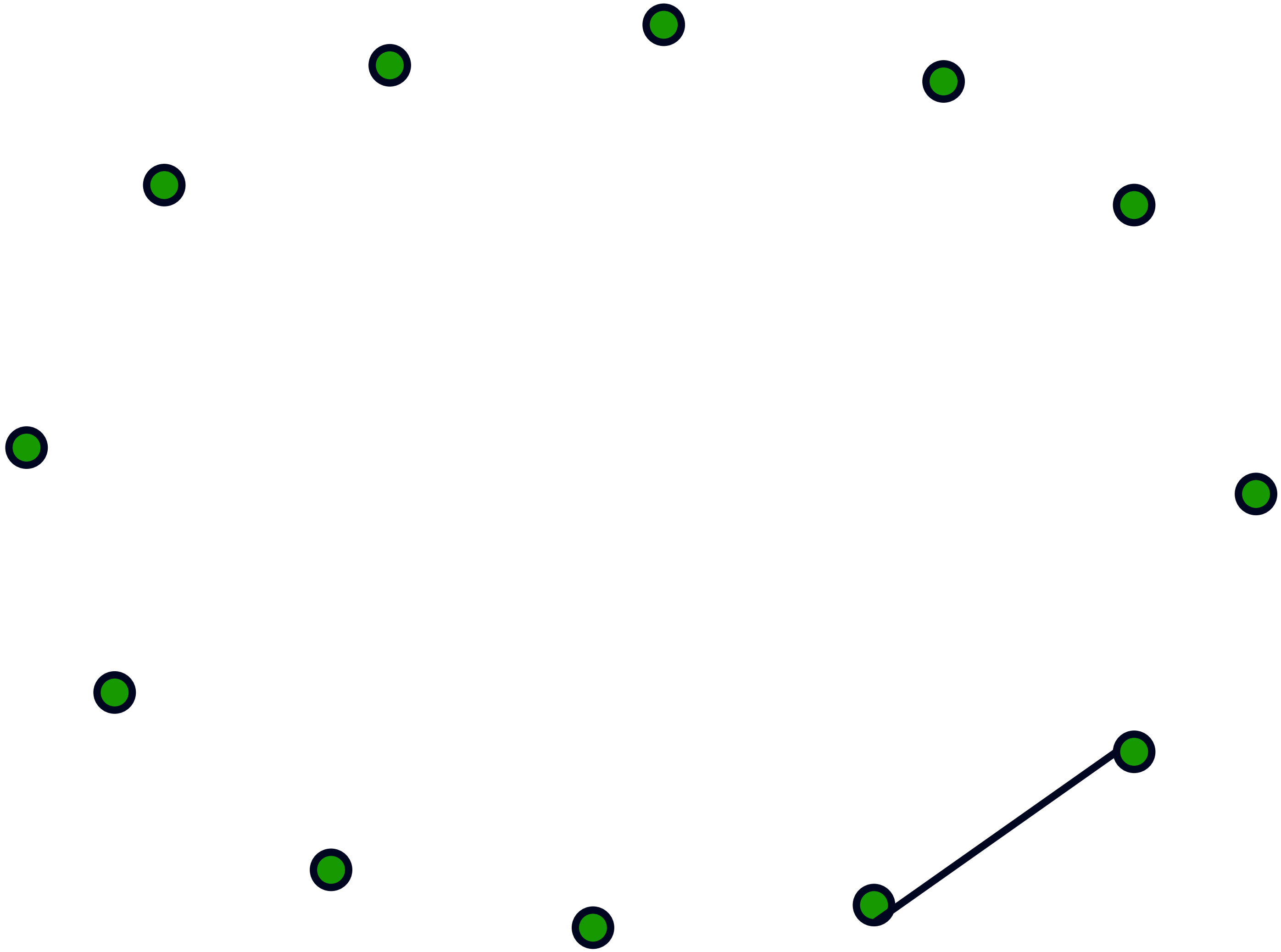
Streaming MaxCut



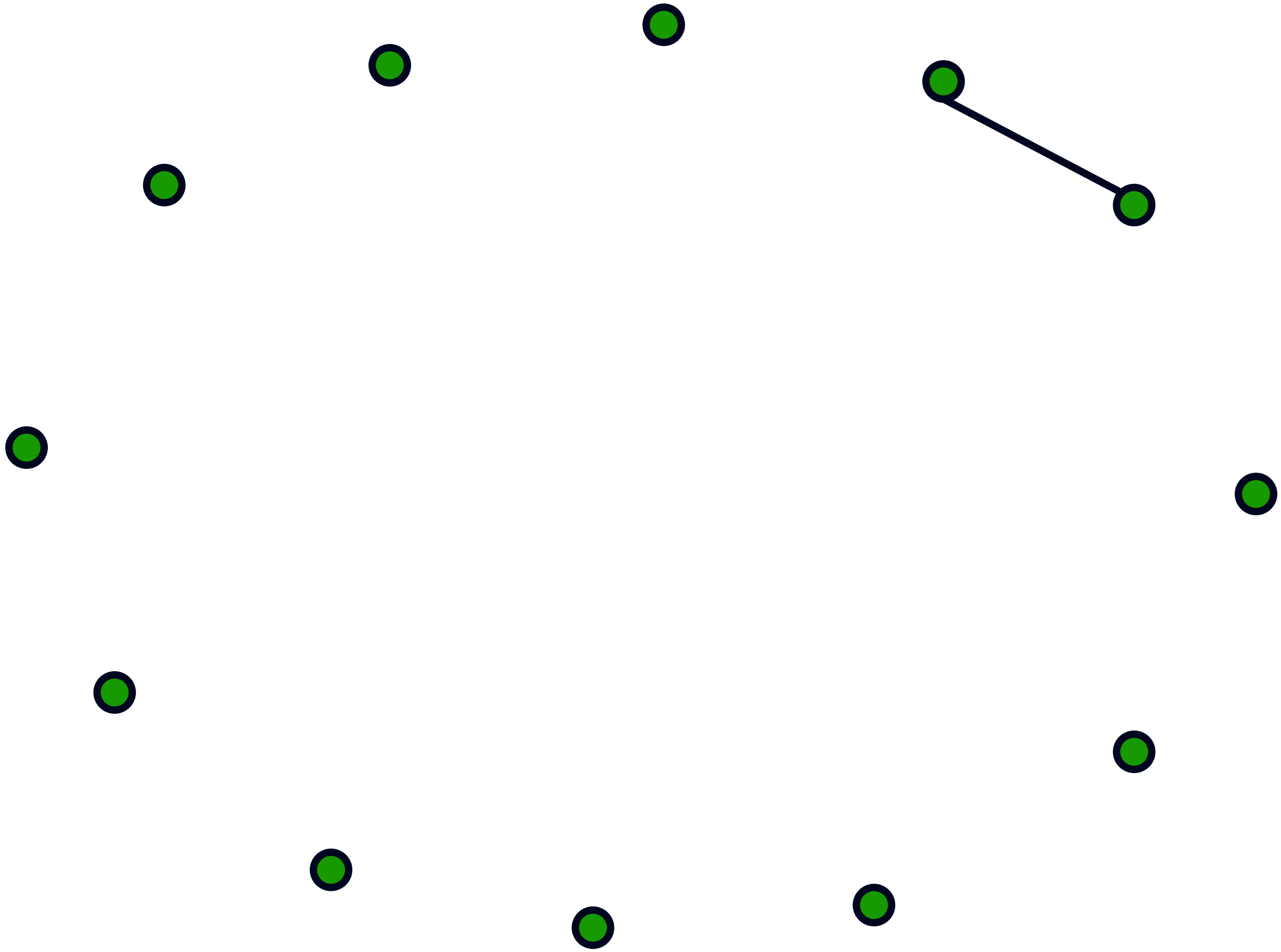
Streaming MaxCut



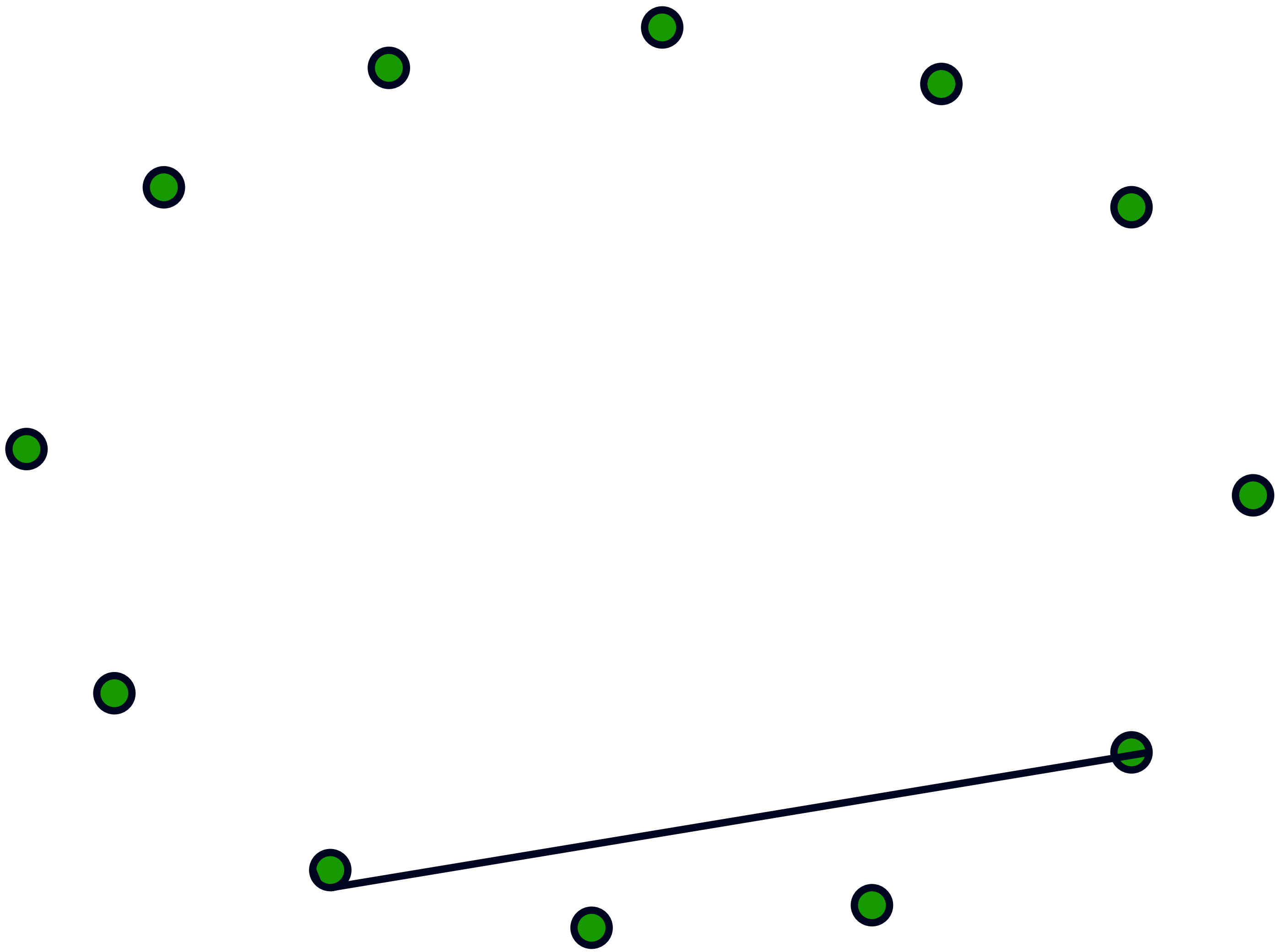
Streaming MaxCut



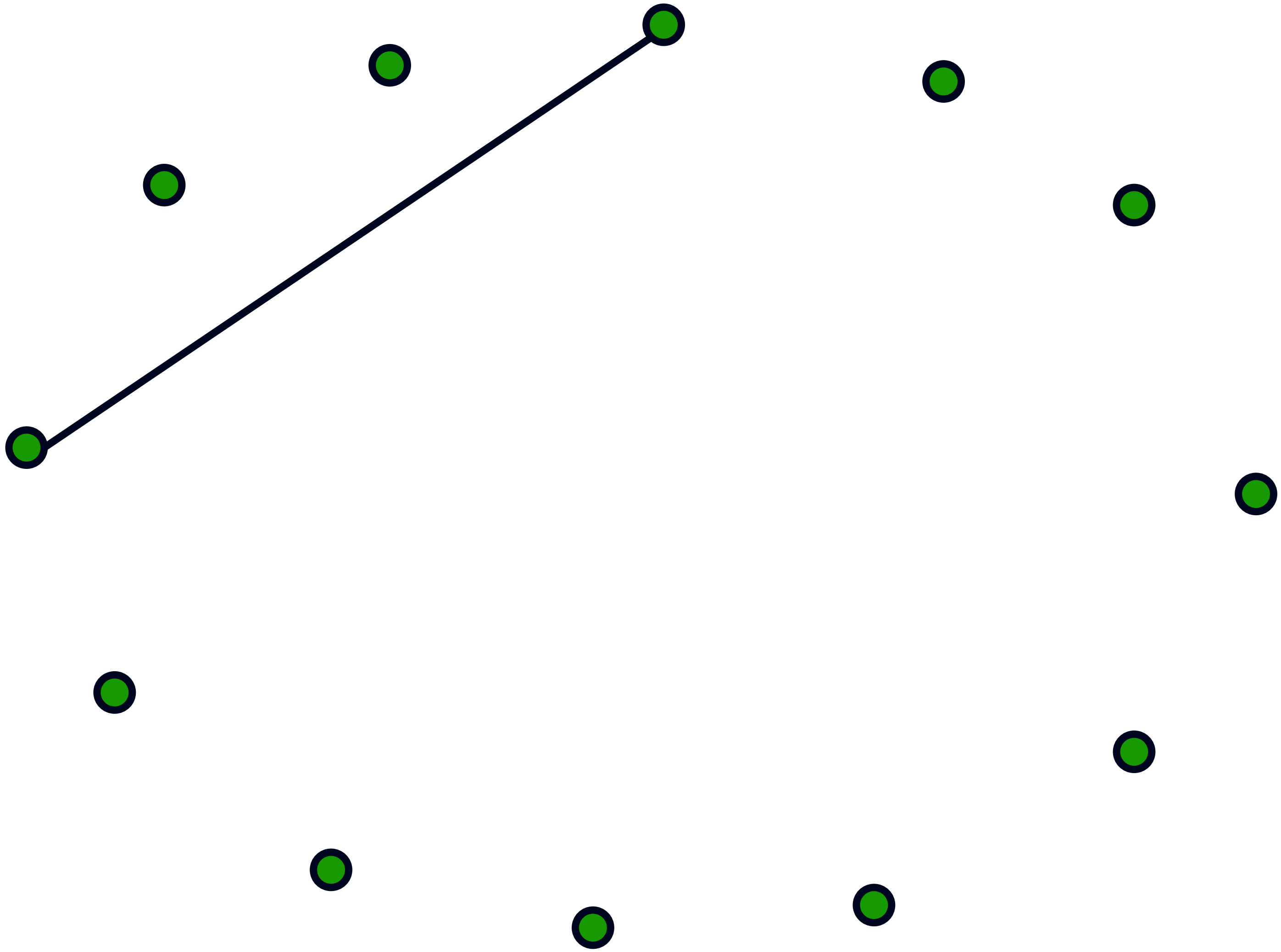
Streaming MaxCut



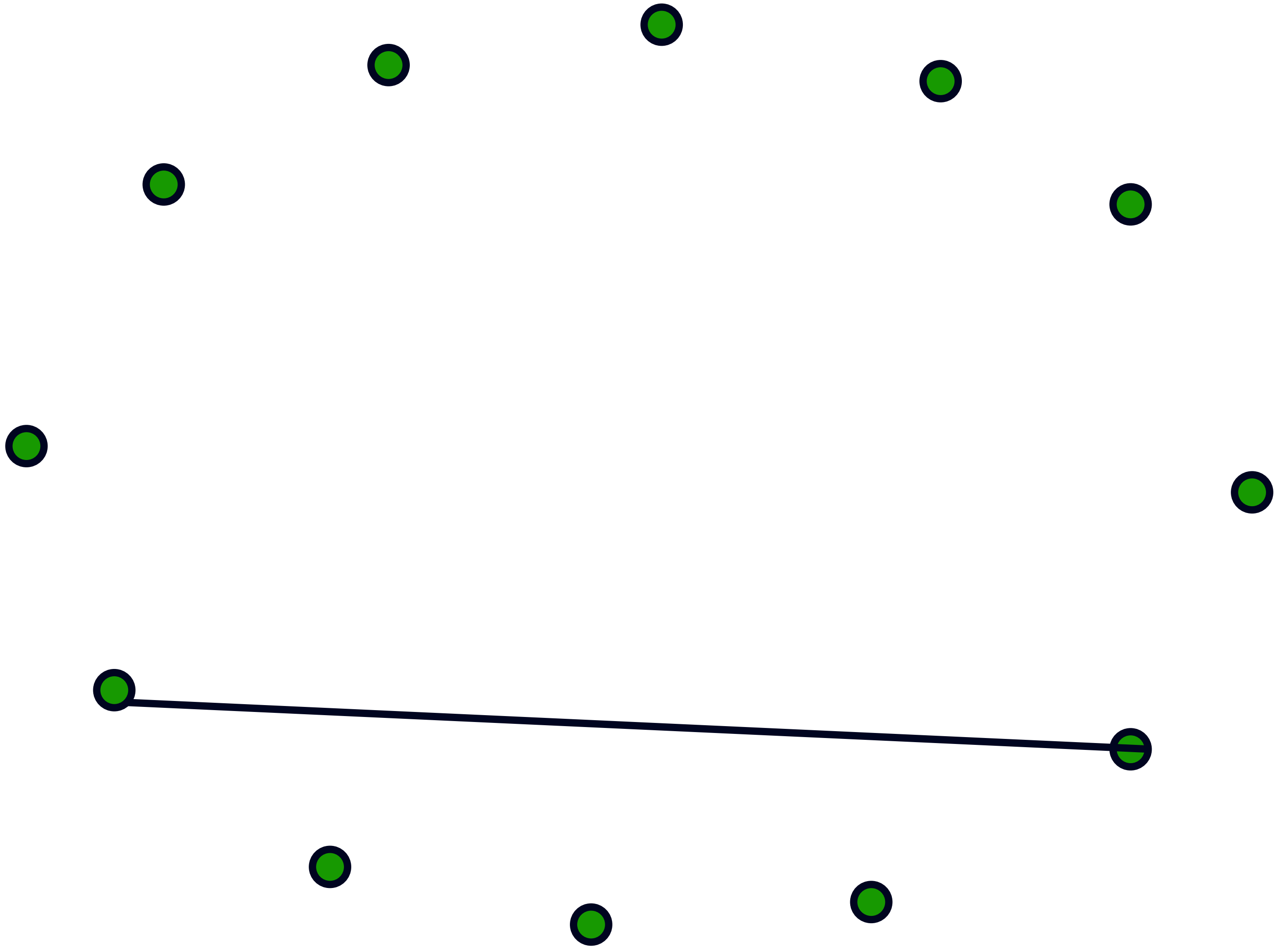
Streaming MaxCut



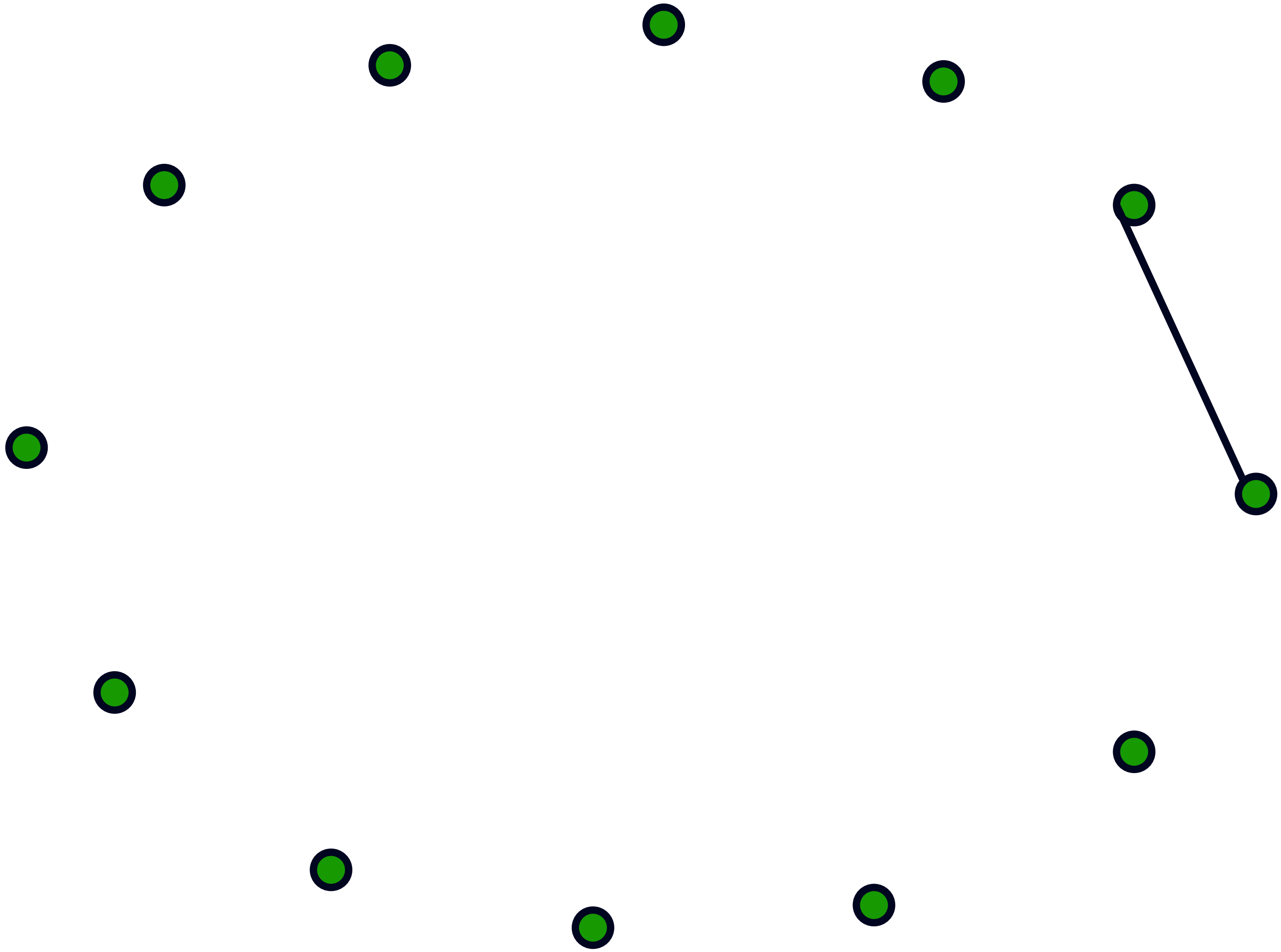
Streaming MaxCut



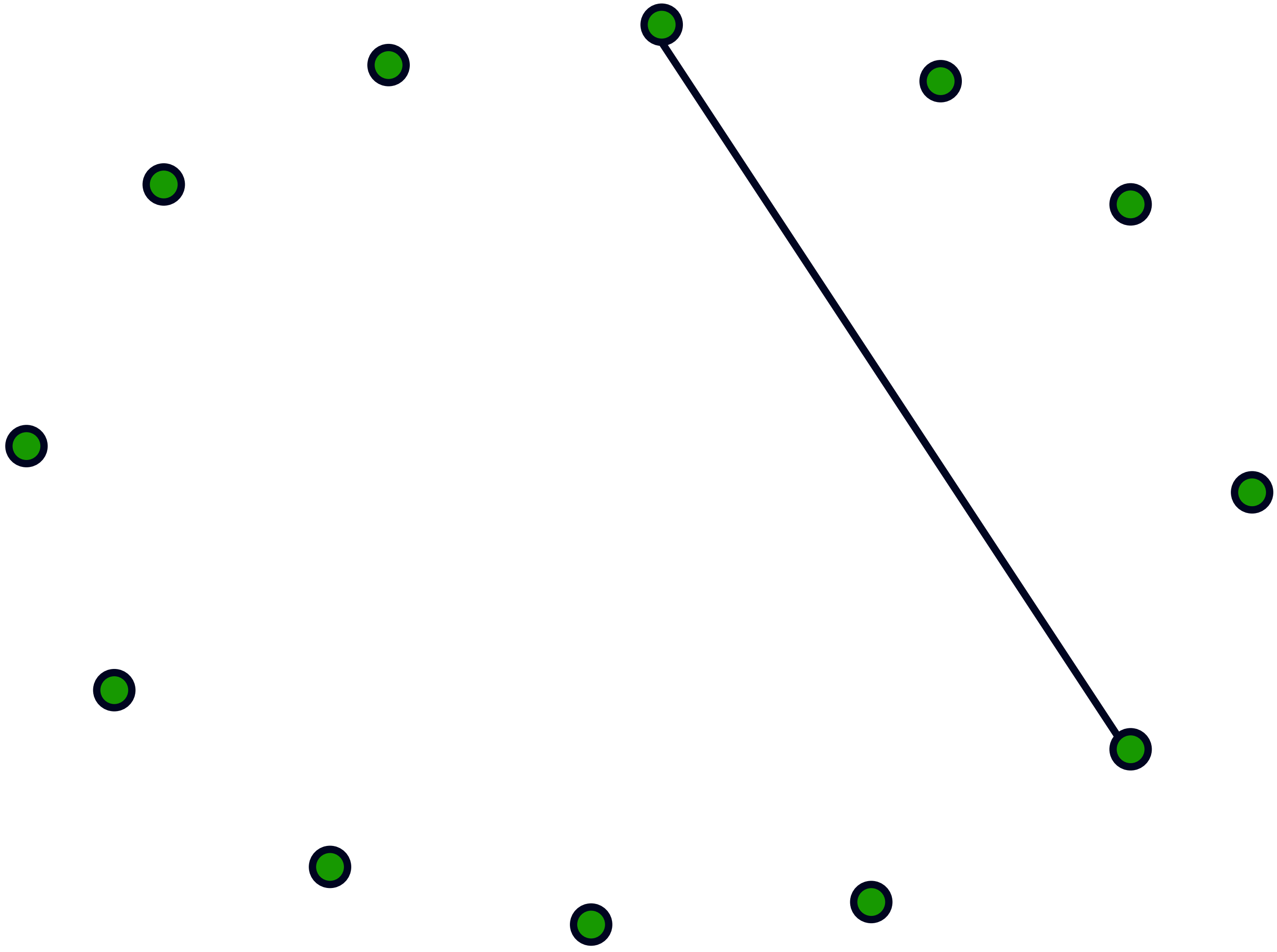
Streaming MaxCut



Streaming MaxCut



Streaming MaxCut



Max Cut Lower Bound

- Hard distributions:
  - YES: Random union of matchings crossing hidden bipartition
  - NO: Random union of matchings.
- Analysis:
  - Divide long stream into O(1) smaller substreams – each substream=matching.
  - Algorithm learns nothing in any single substream $\Leftarrow$ Boolean Hidden Matching Lower Bound [GKKRW]
  - Hybrid argument to combine O(1) substreams.
  - Yields $\Omega(\sqrt{n})$ space lower bound to beat trivial approximation.
- $\Omega(n)$ lower bound more complex – omitted.

Boolean Hidden Matching Problem

- One-way communication problem.
- Alice gets a random cut on vertex set $[n]$.
- Bob gets a random matching on $[n]$ of size αn along with a 0/1 label on each edge.
  - NO: 0/1 labels random
  - YES: 1 ⇒ edge crosses cut, 0 ⇒ doesn't cross.
- Challenge: Alice sends message to Bob, Bob to distinguish YES from NO.
- Lower bound theorem [GKKRW]: $\Omega(\sqrt{n})$ communication required.

GVV+CGV Algorithms for Max DiCut

- Define $Bias(v) \stackrel{\text{def}}{=} indeg(v) - outdeg(v)$
- $\text{Bias}(G) \stackrel{\text{def}}{=} \frac{1}{2}\sum_v |Bias(v)|$
- Claim 1 [GVV]:
  - Bias can be estimated in polylog space. (ℓ_1-norm estimation)
  - $Dicut(G) \leq \frac{Bias(G)+m}{2}$
  - $Bias(G) \leq Dicut(G)$ (Greedy rounding)
  - Output: $\max\left\{\frac{m}{4}, Bias(G)\right\} \Rightarrow$ 2/5-approx.
- Claim 2 [CGV]: $\text{Bias(G)} \leq \frac{m}{3} \Rightarrow Dicut(G) \geq \frac{m}{4} + \frac{Bias(G)^2}{4(m-2Bias(G))}$ (Rand. Rounding w.p. $\frac{1}{2} - \frac{Bias(G)}{2(m-2Bias(G))}$) $\Rightarrow$ 4/9-approx.

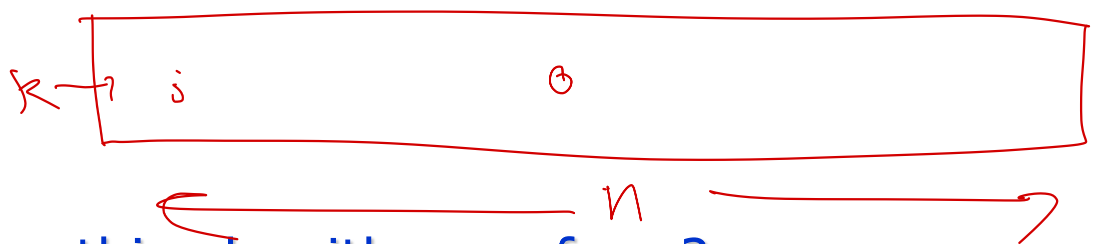
Generalizing to other CSPs: Challenges

- What is bias, say for $f(x, y, z) = x \wedge (y \oplus z)$
  - E.g.: $x \wedge (a_1 \oplus a_2), a_3 \wedge (x \oplus y), y \wedge (x \oplus a_4)$
- Why is ℓ_1-estimation useful in Dicut approximation
  - $\ell_1(x_1, \ldots, x_n) = \max_{a_1 \ldots a_n \in \{-1,1\}} \{\sum_i a_i x_i\}$
  - Useful generalization: $\left|\left|M\right|\right|_{\{1,\infty\}} := \max_{a \in [q]^n} \left\{\sum_i M_{i,a(i)}\right\}$
  - Computable in polylog space
- Dicut Analysis: Graph theory, some 3-var. calculus, Why did rounding end up optimal?

Dichotomy for Sketching

- Stepping back: Suppose algorithm gets entire "incidence matrix" M (and only this info)

  ($M_{ij} =$ fraction of constraints with X_i in jth place in constraint.)



  - How well can this algorithm perform?

Dichotomy for Sketching

- Stepping back: Suppose algorithm gets entire "incidence matrix" M (and only this info)

  ($M_{ij} =$ fraction of constraints with X_i in jth place in constraint.)

  - How well can this algorithm perform?
- Tautology: $\exists \Psi_1, \Psi_2$ with $val(\Psi_1) \geq \gamma, val(\Psi_2) \leq \beta$ and $M(\Psi_1) = M(\Psi_2)$ <u>iff</u> algorithm can't solve (γ, β)-MAX CSP(f).

Dichotomy for Sketching

- Stepping back: Suppose algorithm gets entire "incidence matrix" M (and only this info)
  ($M_{ij} =$ fraction of constraints with X_i in jth place in constraint.)
  - How well can this algorithm perform?
- Tautology: $\exists \Psi_1, \Psi_2$ with $val(\Psi_1) \geq \gamma, val(\Psi_2) \leq \beta$ and $M(\Psi_1) = M(\Psi_2)$ iff algorithm can't solve (γ, β)-MAX CSP(f).
- Thm: - Alg can be sketched in polylog space;
  - If Alg can't solve then no $o(\sqrt{n})$-sketching alg

Dichotomy for Sketching

- Stepping back: Suppose algorithm gets entire "incidence matrix" M (and only this info)
  ($M_{ij} =$ fraction of constraints with X_i in jth place in constraint.)
  - How well can this algorithm perform?
- Tautology: $\exists \Psi_1, \Psi_2$ with $val(\Psi_1) \geq \gamma, val(\Psi_2) \leq \beta$ and $M(\Psi_1) = M(\Psi_2)$ iff algorithm can't solve (γ, β)-MAX CSP(f).
- Thm: - Alg can be sketched in polylog space;
  - If Alg can't solve then no $o(\sqrt{n})$-sketching alg
  - Criterion is decidable in finite time.

Decidability and Criterion (Some ideas)

- Instance only needs to have kq variables.
  - No instances: $\mathrm{val}(\Psi) \leq \beta \Rightarrow \mathrm{val}\big(\wedge_\Pi (\Psi \circ \Pi)\big) \leq \beta$
    - Constraints on first q variables captures Ψ
  - Yes instances: Might as well plant the good assignment!
    - Use var $X_{i\sigma}$ for ith place var assigned $\sigma \in [q]$
- Criterion?
  - Constraints on kq vars $\rightarrow$ Distribution on $[q]^k$
  - Sets $S_Y = \{\Psi \mid \mathrm{val}(\Psi) \geq \gamma\}$ and $S_N = \{\Psi \mid \mathrm{val}(\Psi) \leq \beta\}$ are convex sets!
  - Sets capturing $M(\Psi)$ also convex (in $\mathbb{R}^{kq}$)

Algorithm and lower bound (some ideas)

- If $\{M(\Psi) | \Psi \in S_Y\} \cap \{M(\Psi) | \Psi \in S_N\} = \emptyset$ then there exists a separating hyperplane.
  - Use separating hyperplane to define bias … and get algorithm. (details omitted).
- If $\{M(\Psi) | \Psi \in S_Y\} \cap \{M(\Psi) | \Psi \in S_N\} \neq \emptyset$ then $\exists D_Y, D_N$ on $[q]^k$ with matching marginals.
  - Build a comm. Complexity problem around such a pair D_Y, D_N that extends Boolean Hidden Matching. "(D_Y, D_N)-RMD"
  - Extend the BHM lower bound to all (D_Y, D_N)-RMD (with matching marginals).
  - Use to prove ~~streaming~~ sketching lower bound.

Open Questions

- Sketching = Streaming?
  - Extend sqrt lower bounds from sketching to streaming!
  - Challenge: Walk length algorithm!
- Sqrt = Linear?
  - Is there a dichotomy at linear space?
  - Challenge: Template-based algorithms!
  - Can Dicut approximation be improved with o(n) space?
- $\omega(n)$-lower bounds? (Trivial: $O(n \log n)$ …)

Thank You!