Selected Combinatorial Problems Through the Prism of Random Intersection Graphs Models

Paul G. Spirakis, University of Liverpool, UK

Joint work with:
Sotiris Nikoletseas, University of Patras, Greece
Christoforos Raptopoulos, University of Patras, Greece

13th International Conference on Algorithms and Complexity (CIAC)
Larnaca, Cyprus
June 13-16, 2023

Overview

- Random Intersection Graphs – Basic Model Definition
- Maximum Cliques in RIGs – The Single Label Clique Theorem
- Weighted MAX-CUT in RIGs and Discrepancy in Random Set Systems
- Open problems

Random Intersection Graphs – Basic Model Definition

Introduced in [Karoński et al., Comb. Prob. Computing, 1999]

Definition (Random intersection graph $\mathcal{G}_{n,m,p}$)

- Let $\mathcal{M}$ be a set of m labels.
- Let V be a set of n vertices.
- Edge appearance rule:
  - To each vertex $v \in V$ we independently assign a random label subset S_v, by including each label $\ell \in \mathcal{M}$ independently with probability p.
  - We connect two vertices u, v iff $S_u \cap S_v \neq \emptyset$.

Note 1: The bipartite graph $B_{n,m,p}$ with vertex set $V \bigcup \mathcal{M}$ and edge set $\{(v, \ell) : \ell \in S_v\}$ is called associated bipartite to the $G_{n,m,p}$.

Note 2: Denote by L_ℓ the set of vertices having chosen label ℓ.

An example

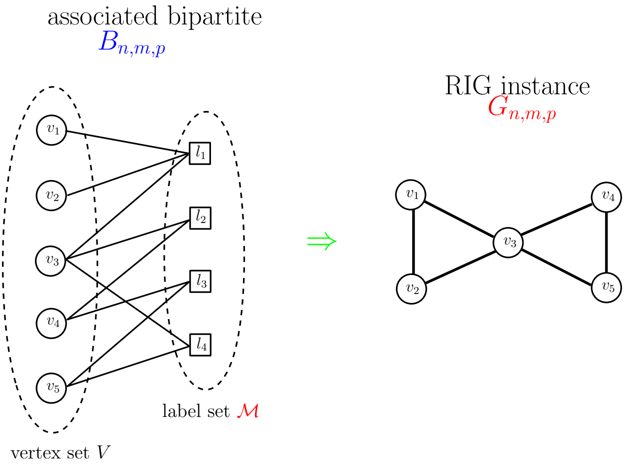


Figure: $S_{v_1} = \{l_1\}$ and $L_{l_4} = \{v_3, v_5\}$

Motivation – Why RIGs

- Appropriate for network modeling in applications where there is an implied dependence between neighboring nodes:
  - social networks (vertices $\leftrightarrow$ individuals, labels $\leftrightarrow$ beliefs/preferences/locations)
  - oblivious resource sharing under various distributed settings
  - secure communication in sensor networks
- Edges are not independent, however...

$\rightarrow$ Equivalence with Erdős-Rényi random graphs: $m = n^{\alpha}, \alpha > 0$ constant

- $\underline{\alpha > 6}$: $\|\mathcal{G}_{n,m,p} - \mathcal{G}_{n,\hat{p}}\|_{TV} \rightarrow 0$, where $\hat{p} = 1 - (1 - p^2)^m$ (uncoditioned edge existence probability) [Fill et al., Rand. Struct. Algorithms, 2000]
- $\underline{\alpha \geq 3}$: equivalence of sharp threshold functions [Rybarczyk, Rand. Struct. Algorithms, 2011]
- $\underline{\alpha > 1}$: translation results regarding lower bounds for increasing properties [Raptopoulos, Spirakis, ISAAC, 2005]

$\rightarrow$ **"interesting case"** $m = n^{\alpha}, \alpha \leq 1$.

Overview

- Random Intersection Graphs – Basic Model Definition
- Maximum Cliques in RIGs – The Single Label Clique Theorem
- Weighted MAX-CUT in RIGs and Discrepancy in Random Set Systems
- Open problems

Definition (MAX-CLIQUE)

Given a graph $G = (V, E)$, find a clique of maximum size.

- Fundamental problem motivated by research related to the Internet, social networks, bibliographic databases, energy distribution networks, global networks of economies etc.

Arbitrary graphs:

- MAX-CLIQUE is NP-complete [Karp 1972]; fastest algorithm runs in $O(1.1888^n)$.
- It is hard to approximate [Håstad 1999]; best approximation ration is $O\left(\frac{n(\log\log n)^2}{(\log n)^3}\right)$.
- Given the maximum clique has size k it cannot be solved in time $n^{o(k)}$, unless the exponential time hypothesis fails; brute force approach takes $O(n^k k^2)$.

Finding Maximum Cliques in Random Graphs

Erdős-Rényi random graphs:

- MAX-CLIQUE remains hard in $G_{n,\frac{1}{2}}$; there is a clique of size $2\ln n$ whp but most algorithms can only find a clique of size $\ln n$.
- Conjecture [Jerrum 1992]: finding an $1.01\ln n$ clique remains hard even if the input graph is a $G_{n,\frac{1}{2}}$ random graph in which we have planted a randomly chosen clique of size $n^{0.49}$.

Random intersection graphs:

- When $m = n^{\alpha}, \alpha < 1$, whp we can find a maximum clique in $G_{n,m,p}$ in polynomial time, provided the label choices are given as input (rather than just the resulting graph). [Nikoletseas, Raptopoulos, Spirakis, Comput. Sci. Rev. 2021], [Nikoletseas, Raptopoulos, Spirakis, MFCS 2012]

Maximum Cliques in RIGs

Label Reconstruction and the Single Label Clique Theorem

Label Reconstruction in $\mathcal{G}_{n,m,p}$

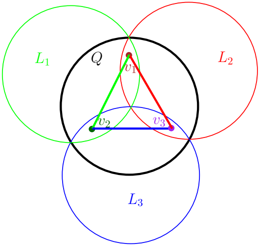
Observation: Given $G_{n,m,p}$, the associated bipartite graph is not unique! But...

Theorem ([Nikoletseas, Raptopoulos, Spirakis, Comput. Sci. Rev. 2021])

Under "mild" conditions on m, p, the bipartite graph $B_{n,m,p}$ associated to $G_{n,m,p}$ is uniquely determined whp, up to permutations of the labels.*

* $m < n,\ p = \Omega\left(\sqrt{\frac{1}{nm}}\right)$ and $mp^2 = O(1)$

Main technical idea: The Single Label Clique Theorem for maximum cliques in "dense" $G_{n,m,p}$



Theorem (Single Label Clique Theorem - SLCT [Comput. Sci. Rev. 2021])

Let $m = n^{\alpha}, 0 < \alpha < 1$ and $mp^2 = O(1)$. Then whp, any clique Q of size $|Q| \sim np$ in $G_{n,m,p}$ is formed by a single label. In particular, the maximum clique is formed by a single label.

Note:

- For $p = o\left(\sqrt{\frac{1}{nm}}\right)$: $G_{n,m,p}$ is chordal whp [Behrisch et al. 2008]
  $\Rightarrow$ finding a maximum clique is easy.
- For $p = \Omega\left(\sqrt{\frac{1}{nm}}\right)$: $G_{n,m,p}$ is not chordal and can be quite dense for $mp^2 = \Theta(1)$.

Maximum cliques in $\mathcal{G}_{n,m,p}$ with $\alpha < 1$

SLCT - proof sketch.

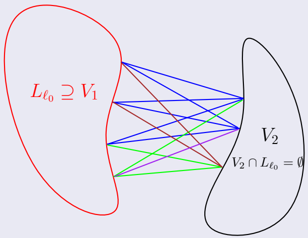
Two definitional tools: for disjoint $V_1, V_2 \subset V$

- $p(V_1, V_2)$: probability all edges between V_1, V_2 exist.

$$p(V_1, V_2) \leq \left(|S^{(2)}_{V_2}|p + \prod_{v \in V_2} \left(1 - (1-p)^{|S_v|}\right) \right)^{|V_1|}$$

  where $S^{(2)}_{V_2}$ is the set of labels chosen by at least 2 vertices in V_2.

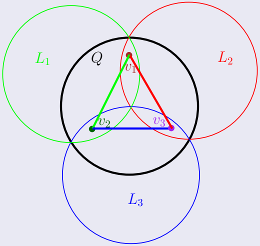
- $A_{|V_1|,|V_2|}$: Event (a) $\exists \ell_0 \in \mathcal{M}$ chosen by all vertices in V_1 but none in V_2 AND (b) all edges between V_1, V_2 exist (we bound the probability of (b) by $p(V_1, V_2)$)



SLCT - proof sketch.

Rainbow coupling (contradiction) argument:

Suppose a large Q is not formed by a single label.



1. For any $\ell_0 \in \mathcal{M}$, we have $|Q \cap L_{\ell_0}| \leq np^{1+c}$, where $0 < c < \frac{1-\alpha}{1+\alpha}$.

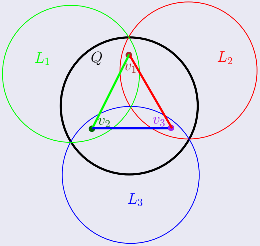
single labels make only small parts of Q

SLCT - proof sketch.

Rainbow coupling (contradiction) argument:

Suppose a large Q is not formed by a single label.

$Q' = \{v_1, v_2, v_3\}$

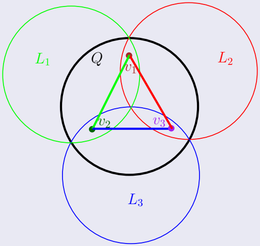


2. By contradiction arguments there is a large rainbow clique $Q' \subseteq Q$:
   - edges in Q' are formed by distinct labels and
   - $|Q'| \geq p^{-\frac{c}{2}}$, for any positive constant $c < \frac{1-\alpha}{1+\alpha}$.

SLCT - proof sketch.

Rainbow coupling (contradiction) argument:

Suppose a large Q is not formed by a single label.



3. By domination and the union bound, the probability that Q' exists is at most

$$\binom{n}{|Q'|}\left(1-(1-p^2)^m\right)^{\binom{|Q'|}{2}} = o(1).$$

□

Is label reconstruction possible?

Corollary

Given the label representation of $G_{n,m,p}$, we can quickly identify its maximum clique whp when $\alpha < 1$.

Open Problem: Can we find $B_{n,m,p}$ (up to permutations of the labels) when provided with just the vertices and edges of $G_{n,m,p}$?

Note: So far this is possible only for small values of m, or for sparse instances of $G_{n,m,p}$.

Overview

- Random Intersection Graphs – Basic Model Definition
- Maximum Cliques in RIGs – The Single Label Clique Theorem
- Weighted MAX-CUT in RIGs and Discrepancy in Random Set Systems
- Open problems

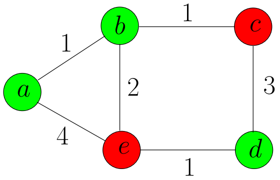
Problem definition: Weighted Max Cut

Definition (Weighted Max Cut)

Input: Undirected weighted graph $G(V, E, \mathbf{W})$, $\mathbf{W}$ is a (symmetric) weight matrix

Output: 2-coloring $\mathbf{x}^{(\max)} : V \to \{+1, -1\}$ with maximum cumulative weight of bicolor edges

$$\max_{\mathbf{x} \in \{\pm 1\}^{|V|}} \frac{1}{4} \sum_{\{i,j\} \in E} \mathbf{W}_{i,j}(x_i - x_j)^2 = \texttt{MaxCut}(G)$$



Problem definition: Weighted Max Cut

Definition (Weighted Max Cut)

Input: Undirected weighted graph $G(V, E, \mathbf{W})$, $\mathbf{W}$ is a (symmetric) weight matrix

Output: 2-coloring $\mathbf{x}^{(\max)} : V \to \{+1, -1\}$ with maximum cumulative weight of bicolor edges

$$\max_{\mathbf{x}\in\{\pm 1\}^{|V|}} \frac{1}{4} \sum_{\{i,j\}\in E} \mathbf{W}_{i,j}(x_i - x_j)^2 = \mathtt{MaxCut}(G)$$

- Well motivated: graph layout and embedding problems, minimizing Hamiltonian of spin models, VLSI design, data clustering etc.
- Complexity: APX-hard (cannot find a solution arbitrarily close to optimal unless P=NP)

Average Case

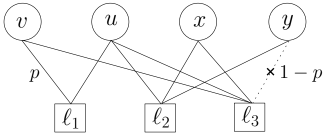
Can we do better when G follows a (known) probability distribution?

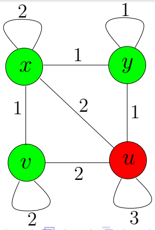
- Sparse Erdős-Rényi random graphs $G_{n,\frac{\gamma}{n}}$, $\gamma > 0$ constant
  - Phase transition at $\gamma = \frac{1}{2}$: $\texttt{MaxCut}(G)$ is $|E| - O(1)$ whp when $\gamma < \frac{1}{2}$, and $|E| - \Omega(n)$ whp when $\gamma > \frac{1}{2}$ [Coppersmith et al., Random Struct. Algorithms,2006]
  - Limiting behaviour for large γ: $\texttt{MaxCut}(G)$ asymptotically equal to $(\frac{\gamma}{2} + P_* \sqrt{\frac{\gamma}{2}})n$, where $P_* \approx 0.7632$, whp. [Dembo et al., Ann. Probability,2017] (non-constructive proof)
  - Approximation algorithm: A cut of size at least $\left(\frac{\gamma}{2} + 0.37613\sqrt{\gamma}\right) n$ whp, can be constructed in polynomial time. [Coppersmith et al., Rand. Struct. Alg.,2004]
- Weighted random intersection graphs (...this talk)

Weighted Random Intersection Graphs – Definition

Definition (Weighted random intersection graph $\overline{\mathcal{G}}_{n,m,p}$)

- Set of n vertices V, set of m labels $\mathcal{M}$
- Assign to each vertex v a random subset $S_v \subseteq \mathcal{M}$: for each $\ell \in \mathcal{M}$, $\Pr(\ell \in S_v) = p$, independently of all other label choices
- Let $\mathbf{R}_{:,v}$ the indicator vector for S_v; $\mathbf{R}$ is the representation matrix of our graph
- Edge $\{u, v\}$ is given weight $|S_v \cap S_u| = [\mathbf{R}^T\mathbf{R}]_{v,u}$
- $G(V, E, \mathbf{R}^T\mathbf{R})$ is a random instance of $\overline{\mathcal{G}}_{n,m,p}$





$\mathbf{R}_{:,v} = (1, 0, 1)^T$, $[\mathbf{R}^T\mathbf{R}]_{v,u} = 2$, $[\mathbf{R}^T\mathbf{R}]_{v,v} = 2$

Further motivation for the weighted model

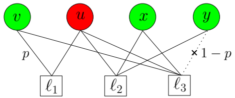
Discrepancy of random set systems

- Let $L_\ell \subseteq V$ the set of vertices having chosen label ℓ in $G(V, E, \mathbf{R}^T\mathbf{R})$.
- Define the set system $\Sigma = \{L_{\ell_1}, L_{\ell_2}, \ldots, L_{\ell_m}\}$.
- $\mathbf{R}$ is then the incidence matrix of Σ; V is the universe for Σ.

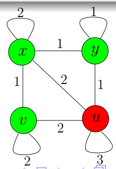
Definition (Discrepancy)

Given a 2-coloring $\mathbf{x} : V \to \{+1, -1\}$

- The imbalance of Σ on $\mathbf{x}$ is measured by
  $\text{disc}(\Sigma, \mathbf{x}) \stackrel{\text{def}}{=} \max_{\ell \in [m]} \left| \sum_{v \in V} \mathbf{R}_{\ell,v} \mathbf{x}_v \right| = \|\mathbf{R}\mathbf{x}\|_\infty$
- The minimum imbalance over all 2-colorings is the discrepancy of Σ
  $\text{disc}(\Sigma) \stackrel{\text{def}}{=} \min_{\mathbf{x} \in \{\pm 1\}^n} \text{disc}(\Sigma, \mathbf{x})$



$\Sigma = \{\{v, u\}, \{u, x, y\}, \{v, u, x\}\}, \qquad \text{disc}(\Sigma) = 1$



Further motivation for the weighted model

Theorem ([Nikoletseas, Raptopoulos, Spirakis, ISAAC 2021])

Let Σ have incidence matrix $\mathbf{R}$ and $G = G(V, E, \mathbf{R}^T\mathbf{R})$. If disc$(\Sigma) \leq 1$, any minimum discrepancy coloring also gives a maximum cut of G and vice versa.

Proof sketch.

Let $\texttt{Cut}(G, \mathbf{x})$ the size of the cut defined by $\mathbf{x}$. By definition

$$\begin{aligned}\texttt{Cut}(G, \mathbf{x}) &= \frac{1}{4}\left(\sum_{i,j \in [n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \|\mathbf{R}\mathbf{x}\|_2^2\right) \\ &\geq \frac{1}{4}\left(\sum_{i,j \in [n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \|\mathbf{R}\mathbf{x}\|_\infty^2\right),\end{aligned}$$

with equality when $\mathbf{R}\mathbf{x}$ has only 0's and ± 1's. □

Further motivation for the weighted model

Theorem ([Nikoletseas, Raptopoulos, Spirakis, ISAAC 2021])

Let Σ have incidence matrix $\mathbf{R}$ and $G = G(V, E, \mathbf{R}^T\mathbf{R})$. If $disc(\Sigma) \leq 1$, any minimum discrepancy coloring also gives a maximum cut of G and vice versa.

Conjecture

When $m = n^{\alpha}, \alpha \leq 1$, there exists a minimum discrepancy coloring $\mathbf{x}^{(disc)}$ that gives a cut $\mathtt{Cut}(G, \mathbf{x}^{(disc)}) = (1 - o(1))\mathtt{MaxCut}(G)$ whp.

- When $n > \frac{1}{\log 4} m \log m, \mathrm{disc}(\Sigma) \leq 1$, whp. [Altschuler,Niles-Weed, Rand. Struct. Algorithms, 2021] (non-constructive proof)

Weighted MAX-CUT in weighted RIGs

Concentration of MAX-CUT for $\alpha \leq 1$

Concentration of Max-Cut for $\alpha \leq 1$

Theorem ([Nikoletseas, Raptopoulos, Spirakis, ISAAC 2021])

Let $G(V, E, \mathbf{R}^T\mathbf{R}) \sim \overline{\mathcal{G}}_{n,m,p}$ *with* $m = n^\alpha, \alpha \leq 1$, *and* $p = \Omega\left(\sqrt{\frac{1}{nm}}\right)$.
Then $\mathit{MaxCut}(G) \approx \mathbb{E}_{\mathbf{R}}[\mathit{MaxCut}(G)]$ *whp with respect to* $\mathbf{R}$.

Note: This means that $\texttt{MaxCut}(G)$ concentrates around its expected value.

Concentration of Max-Cut for $\alpha \leq 1$

Proof sketch.

- Let $\mathbf{x}$ a uniformly random 2-coloring of V.
$$\frac{1}{2}\sum_{i\neq j}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j} \geq \texttt{MaxCut(G)} \geq \mathbb{E}_{\mathbf{x}}[\texttt{Cut}(G,\mathbf{x})] = \frac{1}{4}\sum_{i\neq j}\left[\mathbf{R}^T\mathbf{R}\right]_{i,j}.$$
- By linearity of expectation
$$\mathbb{E}_{\mathbf{R}}[\texttt{MaxCut(G)}] = \Theta(n^2 mp^2).$$
- We will use the following tool:

Theorem (Efron-Stein inequality)

Let $X_1, X_2, \ldots, X_n$ and $X'_1, X'_2, \ldots, X'_n$ be i.i.d. random variables. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{X}^{(i)} = (X_1, \ldots, X_{i-1}X'_i, X_{i+1}, \ldots, X_n)$ (i.e. $\mathbf{X}^{(i)}$ comes from $\mathbf{X}$ by replacing X_i with an independent copy). Then, for any function $f(\mathbf{X})$,

$$Var(f(\mathbf{X})) \leq \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\left(f(\mathbf{X}) - f(\mathbf{X}^{(i)})\right)^2\right]$$

Concentration of Max-Cut for $\alpha \leq 1$

Proof sketch (continued).

- Let $\mathbf{R}^{(\ell,i)}$ be equal to $\mathbf{R}$, but with $\mathbf{R}_{\ell,i}$ replaced by an independent Bernoulli(p) $\mathbf{R}'_{\ell,i}$ $\to$ Define $G^{(\ell,i)} = G(V, E, (\mathbf{R}^{(\ell,i)})^T \mathbf{R}^{(\ell,i)})$.
- $\Pr(\mathbf{R}'_{\ell,i} \neq \mathbf{R}_{\ell,i}) = 2p(1-p)$ and $G, G^{(\ell,i)}$ differ in $\leq |L_\ell \backslash \{i\}|$ edges.
- By Efron-Stein inequality,

$$\begin{aligned} \mathrm{Var}_{\mathbf{R}}(\mathtt{MaxCut}(G)) &\leq \frac{1}{2}\sum_{\ell,i} \mathbb{E}_{\mathbf{R}}\left[\left(\mathtt{MaxCut}(G) - \mathtt{MaxCut}\left(G^{(\ell,i)}\right)\right)^2\right] \\ &= O(n^3 m p^3). \end{aligned}$$

- Concentration follows from Chebyshev's inequality: for any $\epsilon > 0$,

$$\Pr\left(|\mathtt{MaxCut}(G) - \mathbb{E}_{\mathbf{R}}[\mathtt{MaxCut}(G)]| \geq \epsilon n^2 m p^2\right) = O\left(\frac{1}{\epsilon^2 nmp}\right).$$

□

Random Cuts for $\alpha < 1$

Random Cuts for $\alpha < 1$

Theorem ([Nikoletseas, Raptopoulos, Spirakis, ISAAC 2021])

Let $G(V, E, \mathbf{R}^T\mathbf{R}) \sim \overline{\mathcal{G}}_{n,m,p}$ *with* $m = n^\alpha, \alpha < 1$, *and* $p = \Omega\left(\sqrt{\frac{1}{nm}}\right)$.
Let $\mathbf{x}^{(rand)}$ *be a uniformly random 2-coloring of* V.
Whp with respect to $\mathbf{x}^{(rand)}$, $\mathbf{R}$,

$$\mathtt{Cut}(G, \mathbf{x}^{(rand)}) = (1 - o(1))\mathtt{MaxCut}(G).$$

Proof sketch.

- *Recall* $\mathtt{Cut}(G, \mathbf{x}) = \frac{1}{4}\left(\sum_{i,j\in[n]^2} \left[\mathbf{R}^T\mathbf{R}\right]_{i,j} - \|\mathbf{R}\mathbf{x}\|_2^2\right)$.
- *Also* $\mathbb{E}_{\mathbf{R}}[\mathtt{MaxCut}\,(G)] = \Theta(n^2 mp^2)$.

$\rightarrow$ *Suffices to show* $\|\mathbf{R}\mathbf{x}\|_2^2 = o(n^2 mp^2)$ *whp over random* $\mathbf{x}$ *and* $\mathbf{R}$.

Random Cuts for $\alpha < 1$

Proof sketch (continued).

- Define $Z_\ell = \sum_{i \in [n]} \mathbf{R}_{\ell,i} x_i$, i.e. the ℓ-th element of $\mathbf{Rx}$
- Whp, for any ℓ, Z_ℓ has $Y_\ell \le np$ non-zero terms. (Chernoff bound)
- By Hoeffding's inequality, for any $\lambda > 0$

$$\Pr(|Z_\ell| > \lambda | Y_\ell) \le e^{-\frac{\lambda^2}{2Y_\ell}}.$$

- Setting $\lambda = \sqrt{6np \ln n}$,

$$\begin{aligned} \Pr(\exists \ell : |Z_\ell| > \lambda) &\le \Pr(\exists \ell \in [m] : Y_\ell > 3np) + me^{-\frac{\lambda^2}{6np}} \\ &= o(1) + \frac{m}{n} = o(1). \end{aligned}$$

- Overall, whp $\|\mathbf{Rx}\|^2 \le m\lambda^2 = 6nmp \ln n = o(n^2 mp^2)$.

□

Note: Similar proof works also for $n = m$ and $p = \omega(\log n / n)$.

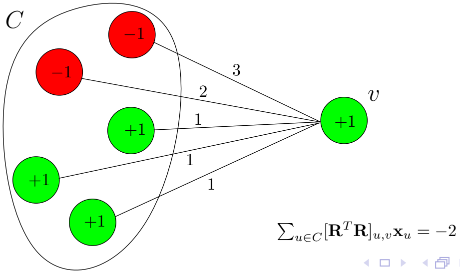
The Majority Cut Algorithm

Large cuts in the symmetric sparse case ($m = n$, $p = \frac{c}{n}$)

Majority Cut Algorithm

Main idea:

1. Colour ϵn vertices at random;
2. Let C the set of coloured vertices;
3. Pick $v \in V \backslash C$ and choose a colour that maximizes the cut in $G[\{v\} \cup C]$, given the colours of C:
   - $\rightarrow$ Coulour $\mathbf{x}_v = -1$ if and only if $\sum_{u \in C} [\mathbf{R}^T \mathbf{R}]_{u,v} \mathbf{x}_u \geq 0$;
4. Set $C = C \cup \{v\}$ and go to step 3;



Main idea:

1. Colour ϵn vertices at random;
2. Let C the set of coloured vertices;
3. Pick $v \in V \setminus C$ and choose a colour that maximizes the cut in $G[\{v\} \cup C]$, given the colours of C:
   - $\rightarrow$ Coulour $\mathbf{x}_v = -1$ if and only if $\sum_{u \in C} [\mathbf{R}^T \mathbf{R}]_{u,v} \mathbf{x}_u \geq 0$;
4. Set $C = C \cup \{v\}$ and go to step 3;

Theorem ([Nikoletseas, Raptopoulos, Spirakis, ISAAC 2021])

Let $G(V, E, \mathbf{R}^T \mathbf{R}) \sim \overline{\mathcal{G}}_{n,m,p}$ with $m = n$, and $p = \frac{c}{n}$, $c > 0$ constant. Whp with respect to $\mathbf{R}$, Majority Cut constructs a cut at least $1 + \beta$ times larger than the expected weight of a random cut, where $\beta \geq 0.43c^{-1.5}$.

Theorem

Let $G(V, E, \mathbf{R}^T\mathbf{R}) \sim \overline{\mathcal{G}}_{n,m,p}$ with $m = n$, and $p = \frac{c}{n}$, $c > 0$ constant. Whp with respect to $\mathbf{R}$, *Majority Cut constructs a cut at least $1 + \beta$ times larger than the expected weight of a random cut, where $\beta \geq 0.43c^{-1.5}$.*

Proof sketch.

- Let M_t the constructed cut size just after the consideration of the t-th vertex:

$$M_t = M_{t-1} + \frac{1}{2}\sum_{i\in[t-1]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,t} + \frac{1}{2}\left|\sum_{i\in[t-1]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,t} x_i\right|.$$

$\rightarrow$ It suffices to bound the expectation of $|Z_t| \stackrel{\text{def}}{=} \left|\sum_{i\in[t-1]} \left[\mathbf{R}^T\mathbf{R}\right]_{i,t} x_i\right|$.

Proof sketch (continued).

- We prove that, for some Binomial random variable Z_t^B,

$$\mathbb{E}[|Z_t||\mathbf{x}_{[t-1]}, \mathbf{R}_{[m],[t-1]}] \geq \mathrm{MD}(Z_t^B),$$

  where $\mathrm{MD}(Z_t^B) \stackrel{\text{def}}{=} \mathbb{E}[|Z_t^B - Z_t'^B|]$, where $Z_t'^B$ is an independent copy of Z_t^B; this is called the mean absolute difference.

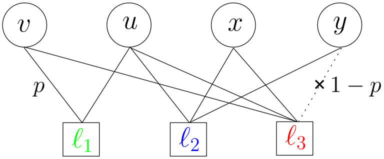
- By the Berry-Esseen Theorem (this is a CLT variant for normal approximation), $Z_t^B - Z_t'^B$ is approximately Normal.
  $\rightarrow$ thus $|Z_t^B - Z_t'^B|$ follows approximately the folded normal distribution:

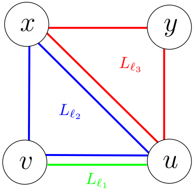
$$\mathrm{MD}(Z_t^B) \geq \sqrt{\frac{c(t-1)}{3\pi n}} - o(1).$$

□

Relating Weighted MAX-CUT in RIGs to
Discrepancy of Random Set Systems

- Note $L_\ell = \{v : \mathbf{R}_{\ell,v}\}$ induces a clique $K^{(\ell)}$.
- We replace $G(V, E, \mathbf{R}^T\mathbf{R})$ with the multigraph $\cup_\ell^+ K^{(\ell)}$; $\cup^+$ means we keep multiplicities.
- Every edge in the multigraph is unweighted (has weight 1).

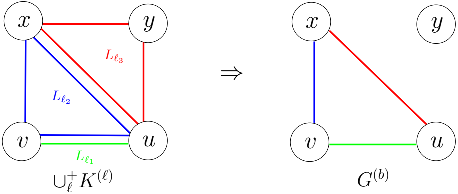




Main idea:

- Try to find a small discrepancy coloring for $\Sigma = \{L_\ell : \ell \in \mathcal{M}\}$ by assigning locally optimal colourings to each L_ℓ separately.
  $\rightarrow$ How can we fix colour inconsistencies?

  1. Replace each clique $K^{(\ell)}$ by a random maximal matching $M^{(\ell)}$, thus getting the graph $G^{(b)} \stackrel{\text{def}}{=} \cup^{+}_{\ell} M^{(\ell)}$;



Main idea:

- Try to find a small discrepancy coloring for $\Sigma = \{L_\ell : \ell \in \mathcal{M}\}$ by assigning locally optimal colourings to each L_ℓ separately.
  $\rightarrow$ How can we fix colour inconsistencies?
  1. Replace each clique $K^{(\ell)}$ by a random maximal matching $M^{(\ell)}$, thus getting the graph $G^{(b)} \stackrel{\text{def}}{=} \cup_\ell^+ M^{(\ell)}$;

Definition (closed vertex-label sequence)

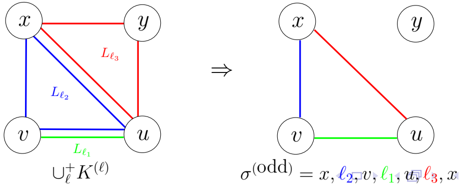
$\sigma := v_1, \ell_1, v_2, \ell_2, \cdots, v_k, \ell_k, v_{k+1} = v_1$ is a closed vertex-label sequence in $G^{(b)}$ if

(i) it has distinct labels and vertices, and

(ii) v_i is connected to v_{i+1} in $G^{(b)}$, for all i.

Main idea:

- Try to find a small discrepancy coloring for $\Sigma = \{L_\ell : \ell \in \mathcal{M}\}$ by assigning locally optimal colourings to each L_ℓ separately.
  $\rightarrow$ How can we fix colour inconsistencies?
  1. Replace each clique $K^{(\ell)}$ by a random maximal matching $M^{(\ell)}$, thus getting the graph $G^{(b)} \stackrel{\text{def}}{=} \cup_\ell^+ M^{(\ell)}$;
  2. Eliminate odd-length closed vertex-label sequences $\sigma^{(\text{odd})}$ as they include odd cycles, thus bipartiteness is violated:
     $\rightarrow$ resample a maximal matching for $K^{(\ell)}$, $\ell \in \sigma^{(\text{odd})}$ ($|L_\ell| \geq 3$);
  3. Repeat step 2 until there are no odd-length vertex-label sequences (if possible);



Maximum cuts in the symmetric sparse case via discrepancy

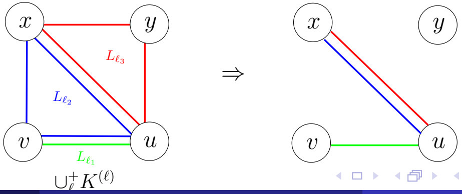
Weak Bipartization Algorithm

Main idea:

- Try to find a small discrepancy coloring for $\Sigma = \{L_\ell : \ell \in \mathcal{M}\}$ by assigning locally optimal colourings to each L_ℓ separately.

  $\rightarrow$ How can we fix colour inconsistencies?

  1. Replace each clique $K^{(\ell)}$ by a random maximal matching $M^{(\ell)}$, thus getting the graph $G^{(b)} \stackrel{\text{def}}{=} \cup^{+}_{\ell} M^{(\ell)}$;
  2. Eliminate odd-length closed vertex-label sequences $\sigma^{(\text{odd})}$ as they include odd cycles, thus bipartiteness is violated:
     $\rightarrow$ resample a maximal matching for $K^{(\ell)}$, $\ell \in \sigma^{(\text{odd})}$ $(|L_\ell| \geq 3)$;
  3. Repeat step 2 until there are no odd-length vertex-label sequences (if possible);



Theorem ([Nikoletseas, Raptopoulos, Spirakis, ISAAC 2021])

Let $G(V, E, \mathbf{R}^T\mathbf{R}) \sim \overline{\mathcal{G}}_{n,m,p}$ with $m = n$, and $p = \frac{c}{n}$, $c > 0$ constant. Whp over $\mathbf{R}$, if Weak Bipartization Algorithm terminates, we can get a minimum discrepancy 2-colouring that also gives a maximum cut of G.

Note: The conditional statement of the above Theorem is not void.

Theorem ([Nikoletseas, Raptopoulos, Spirakis, ISAAC 2021])

Let $G(V, E, \mathbf{R}^T\mathbf{R}) \sim \overline{\mathcal{G}}_{n,m,p}$ with $m = n$, and $p = \frac{c}{n}$, $0 < c < 1$ constant. Whp over $\mathbf{R}$, Weak Bipartization Algorithm terminates in $O\left(\left(n + \sum_\ell |L_\ell|\right) \cdot \log n\right)$ polynomial time.

Conjecture

Let $G(V, E, \mathbf{R}^T\mathbf{R}) \sim \overline{\mathcal{G}}_{n,m,p}$ *with* $m = n$*, and* $p = \frac{c}{n}$, $1 \leq c < e$ *constant. Whp over* $\mathbf{R}$*, Weak Bipartization Algorithm* *terminates in polynomial time.*

Conjecture

When $n = m^\alpha, \alpha \leq 1$*, there exists a minimum discrepancy coloring* $\mathbf{x}^{(disc)}$ *that gives a cut* $\mathtt{Cut}(G, \mathbf{x}^{(disc)})$ *asymptotically equal to* $\mathtt{MaxCut}(G)$ *whp.*

Some more General Research Questions

- Other connections between fundamental combinatorial problems and graph problems on RIGs
- Could algorithmic bottlenecks for the former be translated to bottlenecks for the latter?
- What about temporal RIGs? (definitions, problems, etc.)

Thank you for your attention!