





PLANTED	DENSE SUBMATRIX	
Vertex labels	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$G(\mathbf{n}, \mathbf{k}, \lambda)$
· · · · · · · · · · · · · · ·		· · · · · · · · · · · · · · · · · · ·
	· · · · · · · · · · · · · · · · · · ·	
	· · · · · · · · · · · · · · · · · · ·	
	· · · · · · · · · · · · · · · · · · ·	
	· · · · · · · · · · · · · · · · · · ·	density ~ X
	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·
· · · · · · · · · · · · · · · · · · ·		

PLANTED	DENSE SUBMATRIX	.       .
Vertex labels	$   \overline{O_{v}} = \begin{cases}     1 & w. prob \\     \phi & w. prob \\     h \\     h   \end{cases} $	$G_{\tau}(\mathbf{n}, \mathbf{k}, \lambda)$
Observe Yur	$\sim \begin{cases} \lambda + \mathcal{N}(0, 1) & \sigma_{w} \in \sigma_{v} \in 1 \\ \mathcal{N}(0, 1) & \sigma_{w} \end{pmatrix}$	$\left\{\begin{array}{c} & & \\ & &$
ALGORITHMIC	Q <sub>NS</sub>	density ~ X
<ul><li>Detection :</li><li>Recovery :</li></ul>	determine if whp sample from planted model or all entries N(0,1) given sample from planted model find communities (exactly; weakly	(



PLANTED DENSE SUBMATRIX	- SIMULATIONS
Vertex $O_v = \begin{cases} 1 & w. prob \\ n & n \end{cases}$	$Y_{\mu\nu} \sim \begin{cases} \mathcal{N}(\lambda, 1) & \sigma_{\mu} = \sigma_{\nu} = 1 \\ \mathcal{N}(\alpha, 1) & \sigma_{\mu} = \sigma_{\nu} = 1 \end{cases}$
labels: (p w. prob 1- =	$(\mathcal{I}(\mathcal{O}, \mathbf{I}))$
$H_{1}$ n=100, k=15, $\lambda = 5$	Ho NETUO IAIT TO CONTIN

PLANTED DENSE SUBMATRIX - SIMULATIONS
Vertex $\sigma_v = \begin{cases} 1 & w. prob \\ \emptyset & w. prob \\ w. prob \\  -\frac{k}{n} \end{cases}$ $Y_{uv} \sim \begin{cases} \mathcal{N}(\lambda, 1) & \sigma_u = \sigma_v = 1 \\ \mathcal{N}(0, 1) & D. W. \end{cases}$
$H_1$ n= 100, k= 15, $\lambda = 0.5$ $H_0$ n= 100 (all N(0,1))



PLANTED DENSE SUBMATRIX - SIMULATIONS
Vertex $\sigma_v = \begin{cases} 1 & w. prob \\ \emptyset & w. prob \\ w. prob \\  -\frac{k}{n} \end{cases}$ $Y_{uv} \sim \begin{cases} \mathcal{N}(\lambda, I) & \sigma_u = \sigma_v = 1 \\ \mathcal{N}(0, I) & 0. w. \end{cases}$
$H_1$ n= 100, k= 15, $\lambda = 0.5$ $H_0$ n= 100 (all $N(0,1)$ )

SUM = 100 .35



Sum 2 12.36



PLANTED DENSE SUBMATRIX - SIMULATIONS	· · · · · · · · · · · · ·
Vertex $\sigma_v = \begin{cases} 1 & w. prob \ \frac{k}{n} & V_{uv} \sim \begin{cases} \mathcal{N}(\lambda, I) \\ \mathcal{N}(0, I) \end{cases}$	5u= 0v= 1 D.W.
$H_1 = 100, k = 15, \lambda = 0.5$ $H_0 = 0.100$ (all	$\mathcal{N}(0,1)$



Sum of entries n=100, no planted set, samples=500





CONTEXT Detection 'Easier to detect than recover'.

Recovery





Figure 1: Spiked Matrix Model (planted submatrix with elevated mean).  $H_0$ : random  $n \times n$  matrix with each entry independent with distribution N(0, 1).  $H_1$ :  $n \times n$  matrix with each index in set S independently with probability k/n. Each entry independent with distribution  $N(\lambda, 1)$  if  $i, j \in S$  and with distribution N(0, 1) otherwise.





size of planted clique

## Figure 0: Planted clique.

 $H_0: G(n, \frac{1}{2})$  random graph on n vertices where each edge is present independently with probability 1/2.  $H_1: G(n, k, \frac{1}{2})$ , random graph on n vertices where each vertex is part of 'community' S independently with probability k/n. Each edge ij is present independently either with probability 1 if  $i, j \in S$  or with probability 1/2 otherwise.



## Figure 2: Planted dense subgraph.

 $H_0$ : G(n,q) random graph on n vertices where each edge is present independently with probability q.  $H_1$ : G(n,k,q,s) with s > 0, random graph on n vertices where each vertex is part of 'community' S independently with probability k/n. Each edge ij is present independently either with probability q + s if  $i, j \in S$  or with probability q otherwise.

Planted	Community	$G \sim G(n, p)$	$q', k), K \tilde{\epsilon} \left( \begin{pmatrix} n \\ k \end{pmatrix} \right)$	Aij={Be(p) ijeK (Be(q) ow
n points	<b>.</b>			
		•		
· · · · · · · · ·				· · · · · · · · · · · · · · · ·
· · · · · · · · · ·	• • • • • • •	•		
		•		· · · · · · · · · · · · · · · · ·
· · · · · · · · · ·		•••	• • •	
				· · · · · · · · · · · · · · · · ·
		• • •	• • •	
				· · · · · · · · · · · · · · · · · · ·
· · · · · · · · ·	· · · · · · · · · ·	· · · · · · · · · ·		
· · · · · · · · · ·	· · · · · · · · · · ·	· · · · · · · · · · ·	· · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·
				Fig: Jaming Xu, Duke

Planted Community	$G \sim G(n, p, q', k),  K \in \binom{[n]}{k}$	t <sub>ij</sub> = { Be (ρ) { Be (q)	ijeK ow
n points			
K 'community' nodes	•		
n-k non-community "		· · · · · ·	
· · · · · · · · · · · · · · · · · · ·		· · · · · ·	
· · · · · · · · · · · · · · · · · ·	••••	· · · · · ·	
· · · · · · · · · · · · · · · · · · ·	•	· · · · · ·	

Fig: Janing Xu, Duke



Planted Community	$G \sim G(n, p, q', k),$	$K \stackrel{\text{\tiny "e}}{\in} \binom{[n]}{k}$	tij = {	Ве(р) Ве(q)	ijeK ow	· · ·
	· · · · · · · · · · · · · · · · · · ·		• •			
a napoints a a a a a			• •			
				• • • •	• • • •	• • •
K COMMONITY HOURS	•	• • •				
n-k non-community "		•				
· · · · · · · · · · · · · · · · · · ·		•				
with prob. P		•	• •			
		• • •				
		•••				
			• •			
		• •	• •			
		· · · ·				• • •
		•				
		•				
		•		• • • •	• • • •	• • •
	•	• •				
	•					

Fig: Janing Xu, Duke

Planted Community  $G \sim G(n, p, q, k), \quad K \in \binom{[n]}{k} \quad A_{ij} = \begin{cases} Be(p) & ij \in K \\ Be(q) & ow \end{cases}$ 

n points K 'community' nodes n-k non-community " . . . . . . . . . . ) with prob. P



p>q

Fig: Jawing Xu, Duke

Planted Community	$G \sim G(n, p, q', k),  K \in \binom{[n]}{k}  A_{ij} = \begin{cases} Be(p) & ij \in K \\ Be(q) & DW \end{cases}$	· · · · ·
Process	P > 2	
n points k 'community' nodes n-k 'non-community' " hith prob. p g		.         .         .         .           .         .         .         .         .           .         .         .         .         .           .         .         .         .         .           .         .         .         .         .           .         .         .         .         .           .         .         .         .         .           .         .         .         .         .           .         .         .         .         .           .         .         .         .         .           .         .         .         .         .
Output un labelled graph		· · · · · · · · · · · · · · · · · · ·
		· · · · ·
.       .	Fig: Janing Xu,	Duke

Planted Community	$G \sim G(n, p)$	$q', k), K \in \binom{n}{k}$	$A_{ij} = \begin{cases} Be(f) \\ Be(f) \end{cases}$	p) ijeK
Process	P > 9			) ow
n points				· · · · · · · · · · · ·
<ul> <li>1&lt; 'community nodes</li> <li>n-k 'non-community' "</li> </ul>				
• With prob. P			dr	aw dot it
			> n	in graph
Output P g-				
unlabelled graph L				· · · · · · · · · · · ·
· · · · · · · · · · · · · · · · · · ·				· · · · · · · · · · · ·
· · · · · · · · · · · · · · · · · · ·		arean a guint chaireach. Tha anns an tha seach		· · · · · · · · · · · · ·
· · · · · · · · · · · · · · · · · · ·	n=200 k=50	P=0.3 g=0.1	Fig: J	Faming Xu, Duke

Planted Community	$f= \left( \left( \left( n \right) \right) \right) = \left( \left( n \right) \right) \left( \left( n \right) \right) \right) = \left( \left( n \right) \right) \left( \left( n \right) \right) \right)$	A;; =	SBe (p)	ъjeK		•
Poppage	$\frac{1}{2} = \frac{1}{2} + \frac{1}$		(Be(q)			
	· · · · · · · · · · · · · · · · · · ·	· · · ·				
n points a a a a a						•
1 (community) nodes						
						•
n-c non-community						
		• • • • •				
With prob. P		A CARA				•
• • • • • • • • • • • • • • • • • • •						
		•			• • •	•
		• • • •				
Output						
into bollod acorb						
un abelie a graph						٠
		a a a				•
	a service and the service of the ser					
					• • •	•
· · · · · · · · · · · · · · · · · ·	n=200 k=50 p=0.3 q=0.1		- -			
			g: Jiaw	und Vin i	Duke	•



Planted Community	$G \sim G(n, p, q', k),  K \in \binom{[n]}{k}$	4;; = ∫Be ∕Be	(ρ) <sup>i</sup> jeK	
Process	$\frac{1}{2}$			
n points	M.			
K 'community' nodes		· · ·		
n-k non-community "		• • •		
		• • •	· · · · · · · ·	
With pros. P				
· · · · · · · · · · · · · · · · · · ·				
Output		· · ·		
un labelled graph		· · ·		
		• • •		
· · · · · · · · · · · · · · · · · · ·	n=200 k=50 $p=0.3$ $q=0.1$	· · · ·	T and V	
	· · · · · · · · · · · · · · · · · · ·	. +rg:	Janing Nu,	Duke

Planted	Clique	$G \sim G(n,$	12, K),	$K \stackrel{\text{\tiny e}}{\in} \binom{n}{k}$	$A_{ij} = \begin{cases} A_{ij} = \\ B_{ij} = \end{cases} $	l ijeK 3e(±) ow	
Two	parameter	5	size of	planted	structur	e	
· · · · · · · · · · ·		· · · · · · · · · · ·	sile of	entre	network	 	
$\mathbb{Q}$ . $\mathbb{V}$	Jhen can	we find	planted	clique		· · · · · · ·	· · · · · · ·
· · · · · · · · · · ·	· · · · · · · · · ·		· · · · · · · · · ·	· · · · · · ·	· · · · · · ·	· · · · · · ·	· · · · · ·
	Q	Q				· · · · · · ·	
· · · · · · · · · · ·				{k }		· · · · · · ·	
· · · · · · · · · · ·					· · · · · · ·	· · · · · · ·	· · · · · · ·
· · · · · · · · · · · ·	6		<b>B</b>		· · · · · · ·	· · · · · · ·	· · · · · · ·
· · · · · · · · · · · ·	· · · · · · · · · ·			· · · · · ·	· · · · · · ·		
						tra: Alex	Wein

Planted Clique IMPOSSIBLE: // POSS	$G \sim G(n, \pm, k)$	.), K Ĕ([n] k)	$A_{ij} = \begin{cases} l & ij \in K \\ Be(-1) & ow \end{cases}$	
$2\log_2 n$ k G'~G(n, $\frac{1}{2}$ ): largest	Jn - clique 26032 n (with	$prob \rightarrow 1$ ) => can	<pre> K  ≤ 2.logz n if find "planted" one amongst "bachground" one.</pre>	
			Fig: Alex Wein	

· ·	Plar /	nted 1111		jue	G~G	$(n, \frac{1}{2}, k)$	, $K \stackrel{\text{\tiny "}}{\in} \left( \begin{bmatrix} n7 \\ k \end{bmatrix} \right)  A_{ij} = \begin{cases} \\ \\ \end{cases}$	l ijeK Be(=j) ow
· · ·	· · · · · · · · · · · · · · · · · · ·	210	g <sub>z</sub> n	k.	t Jn	· · · · · · · · · · · · · · · · · · ·	$ f   K  \leq 2 \log_2 n$	
G	' ~ (-	(n,ź	):: lai	rgest c	slique uhp	~ 2 logz n =	> can't find "planted" one in amongst "bachgrow	d' one
M	ethod	s to	find	clique	2			
(	D K	EGREE = set	of K	vertice s	of highest	degree	Can interative version	ר. ר.
• •	Thm [K	(uč 95]	K= 7	L (In log r	⊼) ⇒ P	$(\hat{K} = K) \rightarrow I$	Last I (Un) enoug	, h
• •	· · ·							
• •	• • •	• • •	• • • •					
• •	• • •				· · · · · ·	· · · · · · · · ·	· · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · ·
· ·	· · ·	· · ·	· · · · ·	· · · ·			· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · ·
· · ·	· · · ·	· · · ·	· · · · ·				· · · · · · · · · · · · · · · · · · ·	

Planted Clique G~G(n, ±, K),	$K \stackrel{\text{"}}{\in} \begin{pmatrix} [n] \\ k \end{pmatrix} \qquad A_{ij} = \begin{cases} l & i j \in K \\ Be(\frac{l}{2}) & aw \end{cases}$
$2\log_2 n$ $\sqrt{n}$ $G' \sim G(n, \frac{1}{2})$ : largest clique who ~ $2\log_2 n$ =)	if $ K  \leq 2\log_2 n$ can't find "planted" one
	in amongst "bachground" one
Methods to find clique	· · · · · · · · · · · · · · · · · · ·
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	(an interative version ] get _(Vn) enough]
(i) u top eigenvector of W $U_{ij} = \begin{cases} 2A_{ij} - 1 & i \neq j \\ 0 & 0.W \end{cases}$	$\frac{Thm}{K} [Alon Krivelevich Sudakov '98] \\ k = \Omega(Jn) \implies P(\hat{K} = K) \rightarrow  $
(ii) (threshold) K index vector of K largest $ u_1 $ (iii) (degn-up) $\hat{K} = \{v \in V(G) : e(v, \tilde{K}) \gg \frac{3k}{2}\}$	· · · · · · · · · · · · · · · · · · ·

Planted Clique G~G(n, ±, K), IMPOSSIBLE, IMPOSSIBLE,	$K \stackrel{\text{\tiny theta}}{=} \begin{pmatrix} [n] \\ k \end{pmatrix} = \begin{cases} 1 & ij \in K \\ Be(\frac{1}{2}) & ow \end{cases}$
$2\log_2 n$ $\sqrt{n}$ $G' \sim G(n, \frac{1}{2})$ largest clique whp ~ $2\log_2 n$ => Methods to find clique	if $ K  \leq 2\log_2 n$ can't find "plawted" one in amongst "bachground" one.
$ \begin{array}{c} \textcircled{0}  \overbrace{K = set  of  k \; vertices  of  highest \; degree \\ \overbrace{K = set \; of \; k \; vertices  of \; highest \; degree \\ \overbrace{hm \left[Ku\check{k} \; 95\right] \; k = \mathcal{L}\left(\sqrt{n \log n}\right) \implies P(\widehat{k} = K) \longrightarrow 1. \end{aligned} $	[an interative version] [ (Un) enough ]
(i) U top eigenvector of W (ii) U top eigenvector of W (iii) (threshold) $\tilde{K}$ index vector of K largest $ U_1 $ (iii) (clean-up) $\tilde{K} = \{v \in V(G)\} = e(v, \tilde{K}) \gg \frac{3k}{4}\}$	$\frac{Thm}{L=\Omega(\sqrt{n})} \Rightarrow P(\hat{K}=K) \rightarrow  $
3 SDP METHOD Yes. If $k = -\Omega(\sqrt{n})$ .	

Planted Clique $G \sim G(n, \pm, k)$ ,	$K \stackrel{\text{\tiny L}}{\in} \begin{pmatrix} [n] \\ k \end{pmatrix} \qquad A_{ij} = \begin{cases} l & ij \in K \\ Be(=) & aw \end{cases}$
INPOSSIBLE / HARD? / EASY / X	- reductions (aug.case) poly-time alg).
2log2n K.	- restricted class of alg low deg poly
G'~G(n, ±): largest clique whp ~ 2log2n. =>	can't find "planted" one
Methods to find clique	
$ \begin{array}{c} \textcircled{0}  \overbrace{\mathcal{K}}^{DEGREE  Test} \\ \widehat{\mathcal{K}} = set \ \text{of}  \kappa \ vertices  \text{of}  highest \ degree} \\ \hline{\mathcal{Thm}} \left[ Ku\check{x} \ 95 \right]  \kappa = \mathcal{J} \left( \sqrt{n \log n} \right) \implies P(\widehat{K} = K) \longrightarrow 1. \end{aligned} $	(an interative version ] get _A (Vn) enough]
(2) SPECTRAL METHOD $W_{ij} = \begin{cases} 2A_{ij} - 1 & i \neq j \\ 0 & 0.W \end{cases}$	$\frac{Thm}{K} [Alon Krivelevich Sudakov '98] \\ k = \Omega(Jn) \implies P(\hat{K} = K) \rightarrow  $
(ii) (clean-up) $\hat{K} = \{ \forall \in V(G) : e(v, \tilde{K}) > \frac{3k}{4} \}$	
3 SDP METHOD Yes. If $k = -\Lambda(\sqrt{n})$ .	

CONTEXT		PLANTED	CLIQUE	· · · · · · · · · · ·	 	
	Detection			Recovery	 	
· · · · · B'	$k = \Theta(n^{B})$			$k = \Theta(n^{\beta})$	 	
· · · · · · · · ·			· · · · · · · · · · · ·		 	
KACN					 	
CAST .			EASY.		 	
2			1/2		 	
HARD			HARD		 	
INPOSCIPIE					 	
			(MPOSSIKLE)	<b>7</b>	 	
 <b>1</b> . <b>.</b> .					 	
. Ho			recover		 	
· ····H, ·	I I I K I K K K				 	
· · · · ·					 	

CONTERT	PLANTED	CLIQUE								
	$k = \Theta(n^{\beta})$	· · · · · · · · · ·	B	k = 0	$(n^{\beta})$	· · ·	• • •	• •	• • •	· · ·
· · · · · · ·	Δ	· · · · · · · · · ·		· · · · · · ·		· · ·	• • •	• •	• • •	
EASY	bioggor planted	EASY				· · · ·	· · · ·	· ·		· · · ·
	structure	50M2 \	1/2	• • • • • • •	· · · ·	· · ·	 	· ·	· · ·	 
HARD					· · · ·	· · · ·	· · · ·	• •	· · ·	· · ·
IMPOSSIBLE		(MPOSS)	IBIE		· · ·	· · ·		<b>)</b> 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		· · ·
Но		recov	او (	1 1 1 1	· · ·	· · · ·	· · ·	· ·	· · ·	· · ·
٧s . H,		· · · · · · · · · · ·	· · · ·		· · · ·	· · · ·	· · · ·	· · ·	· · ·	· · · ·
	· · · · · · · · · · · · · · · · · · ·							• •		



## Figure 2: Planted dense subgraph.

 $H_0$ : G(n,q) random graph on n vertices where each edge is present independently with probability q.  $H_1$ : G(n,k,q,s) with s > 0, random graph on n vertices where each vertex is part of 'community' S independently with probability k/n. Each edge ij is present independently either with probability q + s if  $i, j \in S$  or with probability q otherwise.

Hypothesis Testing Guen Sample Which model was a	t generated from.
$H_{0}: G \sim P_{n} \cdot G(n, \frac{1}{2})$ $H_{1}: G \sim Q_{n} \cdot G(n, \frac{1}{2}, K)$ $\blacksquare$ $distributions on \mathbb{R}^{\binom{n}{2}}$	
f detects f d	loesn't detect
$k = Var f(6) \rightarrow j$ $k = Var f(6) \rightarrow j$	
() f(G) () () () () () () () () () () () () ()	(1) $f(G)$ $f(G)$
E[f(G)] E[f(G)]	E[f(G)] = [f(G)]
seq_d poly A degree D test fn: R <sup>n2</sup> → R deg ≤ D. strongly seperate	NB: D~log n consider small/fast
Ep[f] - Ep[f] >> Jmax & Varg [f], Varp[f]}	D is log n consider high deg/ slow.

**Further particulars** The course will comprise  $\sim 15$  lectures and  $\sim 5$  problems sessions. The assessment, all of which can be done in small groups (up to 2-3), will be exercise sheets (2×25%) and 1 longer project (50%). The first exercise sheet will be out Friday 3rd and due Monday 21st February, the second will be out Friday 24th March and due 17th April.

For the longer project is to understand the proof of tractability, hardness or impossibility of a particular problem. List of suggestions will be provided (by 21st April) including some reductions in total variation from a paper by Brennan and Breser, spectral method to achieve the threshold in stochastic block from a paper by Lelarge, Bordenave and Massoulié as well as some candidate lemmas which together will prove some new results (probably a new testing problem where both  $H_0$  and  $H_1$  consist of different planted structures instead of planted and null: with lemmas to prove low-deg hardness, find fast algorithms, info-theoretic thresholds). Hand in either ~5-10 pages give or 25 minutes talk each person end of May / early June.

**Dates (provisional)** Lectures and problem sessions all in 64119 unless otherwise indicated, and will start 15min past the hour.

L1 Thu 26th Jan 3-5pm L2 Wed 1st Feb 3-5pm L3 Thur 9th Feb 3-5pm L4 Wed 15th Feb 3-5pm L5 Wed 22nd Feb 3-5pm L6 Wed 1st Mar 3-5pm L7 Wed 8th Mar 3-5pm

4 degree D test fri R <sup>h2</sup> -> R deg & D. strongly sepercites if	$G(n, q; \lambda, M)$
$\begin{split} & E_{p_n}[f] - E_{Q_n}[f] \gg \int max \left\{ Var_{Q_n}[f], Var_{P_n}[f] \right\} \\ & "difference in means" \gg "fluctvations" \\ & F SL ( ) \\ & THM \\ & Given parameters n, k, \lambda, M^{c} \\ \end{split}$	$\langle \cdot \rangle$
$P_{n} \sim G(n, k, \lambda, M)$ $\mathbb{Q}_{n} \sim G(n, k, \lambda, 1)$ $\mathbb{B}$ $(ounting)$ $\mathbb{B}$	$k = \Theta(n^{\beta})$ easy degree 1 test
$D^{5} \lambda^{2} M^{2} \left( \frac{k^{2}}{n} \sqrt{1} \right) = o(i) \implies \underline{N} \circ \operatorname{deg} D  \text{test}$ $Weakly \ separates \ P_{n}, Q_{n}$ $M^{2} \lambda^{2} k^{2} \qquad (2 - 2)  D  (1 - \log k) + \frac{1}{2} k^{2} k^{2}$	hard / impossible no low degree test
$M \land K = w(i) = )$ Deg 1 test which $M$ strongly separates $R_{i}, Q_{i}$ $8 \ k = w(i)$	$\lambda = \Theta(n^{-\alpha})$ $1/2 \qquad 1 \qquad \alpha$