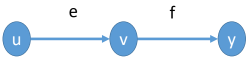
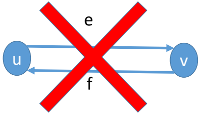
Doodle pour choix des articles à présenter:
https://doodle.com/poll/uis8bxkassgye7y9?utm_source=pollutm_medium=link

Non-backtracking matrix

B: $2m \times 2m$ matrix where m: number of edges of G, defined as $B_{i\to j,k\to\ell} = \mathbb{I}_{j=k}\mathbb{I}_{i\neq\ell}$.





Allows counting of non-backtracking paths in G: $(B^t)_{i\to j,k\to\ell} = \cdots$
$\cdots |\{\text{NB paths with } t+1 \text{ edges, started at } i\to j, \text{ ending at } k\to\ell\}|$.

Spectrum of B: $\lambda_1(B) \geq |\lambda_2(B)| \geq \cdots \geq |\lambda_{2m}(B)|$.

Spectrum of NBM B for sparse SBM $G \sim \mathcal{G}(n, P, \alpha)$

Mean progeny matrix $M = \alpha P$, spectrum:
$\lambda_1(M) = \alpha \geq |\lambda_2(M)| = \alpha|\lambda_2(P)| \geq \cdots \geq |\lambda_q(M)| = \alpha|\lambda_q(P)|$.

Let $x_i \in \mathbb{R}^q$: eigenvector of M associated with $\lambda_i(M)$.
For $e = u \to v \in \vec{E}$, define $y_i(e) = x_i(\sigma_u)$.
For $\ell = c\ln(n)$, $c > 0$ fixed constant, let $z_i = B^\ell B^{\top\ell} y_i$.

Theorem

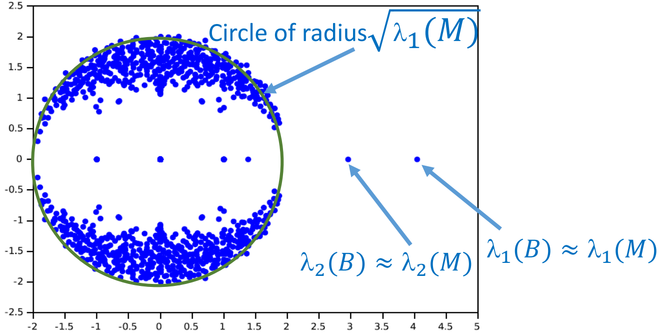
Let $r_0 = \sup\{i \in [q] : \lambda_i(M)^2 > \lambda_1(M)\}$.
(Note: $r_0 \geq 2 \Leftrightarrow \alpha\lambda_2(P)^2 > 1$, *i.e. above Kesten-Stigum threshold).*
Then $\forall i \in [r_0]$, *eigenpair* $(\lambda_i(B), \xi_i)$ *of* B *verifies:*

$\lambda_i(B) \underset{n\to\infty}{\overset{proba.}{\to}} \lambda_i(M)$.

$\exists x_i \in \mathbb{R}^q$*: eigenvector of* $M \leftrightarrow \lambda_i(M)$ *such that for associated* $z_i \in \mathbb{R}^{2m}$,
$\lim_{n\to\infty} \frac{\langle z_i, \xi_i\rangle}{\|z_i\|, \|\xi_i\|} = 1$.

For $i > r_0$, $|\lambda_i(B)| \leq \sqrt{\lambda_1(M)} + o(1)$.

Spectrum of NBM fo $q = 2$, above Kesten-Stigum threshold



Intuition for form of eigenvectors:

For large ℓ,

$$\begin{aligned} z_i := B^{\top\ell} y_i(u \to v) &= \sum_{(u',v')} x_i(\sigma_{v'}) B^{\top\ell}(u \to v, u' \to v') \\ &\approx \sum_{(u',v')} x_i(\sigma_{v'}) \mathbb{I}_{\exists \text{ length } (\ell+1)\text{-NBP } (u' \to v') \longrightarrow (u \to v)} \\ &\approx [\alpha\lambda_i(P)]^\ell Z_i(u), \end{aligned}$$

where $Z_i(u)$: martingale limit as in analysis of Census reconstruction for tree model.

Then by construction $B^\top z_i \approx \alpha\lambda_i(P) z_i$, i.e. z_i: approximate eigenvector.

Relation to Ramanujan graphs

Definition (Lubotzky, Phillips and Sarnak, 1995)

Ramanujan graphs = d-regular graphs with adjacency matrix A such that
$$\sup_{\lambda\in \mathrm{Sp}(A),|\lambda|\neq d} |\lambda| \leq 2\sqrt{d-1}.$$

Recall Alon-Boppana's inequality, $\lambda_2(A) \geq 2\sqrt{d-1}(1-O(\Delta^{-2}))$ where Δ: graph diameter. Hence Ramanujan graphs: regular graphs with optimal spectral gap.

Theorem (Ihara-Bass formula; see lecture notes)

For graph with n vertices, m edges, adjacency matrix $A\in\{0,1\}^{n\times n}$, non-backtracking matrix $B\in\{0,1\}^{2m\times 2m}$, matrix $Q = Diag(\{d_i-1\})\in\mathbb{R}^{n\times n}$, where d_i : degree of node i, then $\forall u\in\mathbb{C},\ (1-u^2)^{n-m}Det(I-uB) = Det(I-uA+u^2Q)$.

Relation to Ramanujan graphs

Corollary

A d-regular graph, $d \geq 2$ is Ramanujan iff its non-backtracking matrix B is such that all eigenvalues λ of B satisfy either $|\lambda| = d - 1 = \rho(B)$, or $|\lambda| \leq \sqrt{\rho(B)}$.

Definition (extended)

Ramanujan graphs: not necessarily regular graphs with NBM B such that for $\lambda \in Sp(B)$, either $|\lambda| = \rho(B)$, or $|\lambda| \leq \sqrt{\rho(B)}$.

Theorem's result implies: for $G \sim \mathcal{G}(n, \alpha/n)$ Erdős-Rényi graph, its NBM has w.h.p. spectrum: $\rho(B) = \alpha + o(1)$, and all other eigenvalues $|\lambda| \leq \sqrt{\alpha} + o(1)$. Hence up to $o(1)$ error, Ramanujan according to extended definition.

Result is a (non-regular) counterpart of [Friedman 2008]: for $d \geq 3$, random d-regular graphs such that w.h.p., $\sup_{\lambda \in Sp(A), |\lambda| \neq d} |\lambda| \leq 2\sqrt{d-1} + o(1)$.

Existence of hard phase [Banks et al. 2016]

For symmetric Potts model with q blocks, parametrized by $\alpha P_{ii} = c_{in}$, $\alpha P_{ij} = c_{out}$, $j \neq i \in [q]$,
average degree $\alpha = \frac{c_{in}+(q-1)c_{out}}{q} = \lambda_1(M)$, $\lambda_2(M) = \frac{c_{in}-c_{out}}{q}$.

Definition

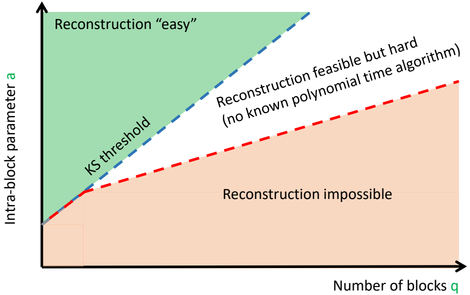
Partition of $[n]$ is **good** if it splits $[n]$ into q equal-size groups, such that the number m_{in} (resp. m_{out}) of edges intra-groups (resp. inter-groups) verifies $|m_{in} - \bar{m}_{in}| \leq n^{2/3}$, $|m_{out} - \bar{m}_{out}| \leq n^{2/3}$,
where $\bar{m}_{in} = \frac{nc_{in}}{2q}$, $\bar{m}_{out} = \frac{n(q-1)c_{out}}{2q}$.

Then: for $q \geq 4$, can find parameters c_{in}, c_{out} such that: $\lambda_2^2 < \lambda_1$, i.e. below Kesten-Stigum threshold, and for some $\epsilon > 0$,
$\lim_{n\to\infty} \mathbb{E}|\{\text{ good partitions with overlap } \leq \epsilon\}| = 0$. Since there exists a good partition w.h.p. (partition $[q]$ according to true spin values σ), this implies that partial reconstruction is feasible below K-S threshold.

Remark

Partial reconstruction of communities in model $\mathcal{G}(n, \alpha, P)$ is polynomial-time feasible above KS threshold, and believed to be infeasible in poly-time below KS threshold.
Hence existence of two or three phases: reconstruction

- *Infeasible*
- *feasible (information-theoretically) but computationally hard*
- *poly-time feasible (above KS threshold)*



Hypothesis testing problems

Example: for graph G, decide whether
$H_0: \ G \sim \mathbb{P}_n: \ \mathcal{G}(n, \alpha/n)$ (Erdős-Rényi graph), or
$H_1: \ G \sim \mathbb{Q}_n: \ \mathcal{G}(n, \alpha, P)$ (sparse SBM).
Let Likelihood ratio $Y_n := \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$. By Neyman-Pearson's lemma, for all $\epsilon > 0$, $\exists$ test $T \in \{0, 1\}$ maximizing $\mathbb{Q}_n(T = 1)$ among tests such that $\mathbb{P}_n(T = 1) \leq \epsilon$ given by

$$\begin{array}{ll} T = 1 & \text{if } Y_n > t, \\ T = 1(0) \text{ with prob. } p(1-p) & \text{if } Y_n = t, \\ T = 0 & \text{if } Y_n < t. \end{array}$$

Definition

Detection between H_0, H_1 (i.e. $\mathbb{P}_n, \mathbb{Q}_n$) is feasible if $\exists$ tests $\{T_n\}_{n>0}$ such that $\mathbb{P}_n(T_n = 1) \overset{n\to\infty}{\to} 0$ and $\mathbb{Q}_n(T_n = 0) \overset{n\to\infty}{\to} 0$.

Definition (Contiguity)

Sequence $\{\mathbb{P}_n\}_{n>0}$ contiguous with respect to sequence $\{\mathbb{Q}_n\}_{n>0}$ iff for all sequence of events $\{E_n\}_{n>0}$,
$\lim_{n\to\infty} \mathbb{Q}_n(E_n) = 0 \Rightarrow \lim_{n\to\infty} \mathbb{P}_n(E_n) = 0$

Lemma

If for some $c > 0$, $\sup_{n>0} \mathbb{E}_{\mathbb{Q}_n} Y_n^2 \leq c$, *then* $\{\mathbb{P}_n\}_{n>0}$ *contiguous with respect to* $\{\mathbb{Q}_n\}_{n>0}$.

Proof: $\mathbb{P}_n(E_n) = \mathbb{E}_{\mathbb{Q}_n}[Y_n \mathbb{I}_{E_n}] \leq \sqrt{\mathbb{Q}_n(E_n)\mathbb{E}_{\mathbb{Q}_n} Y_n^2} \leq \sqrt{c}\sqrt{\mathbb{Q}_n(E_n)}$.

Property: If contiguity (of $\{\mathbb{P}_n\}_{n>0}$ w.r.t. $\{\mathbb{Q}_n\}_{n>0}$, or of $\{\mathbb{Q}_n\}_{n>0}$ w.r.t. $\{\mathbb{P}_n\}_{n>0}$) then detection is infeasible.

Indeed, let $E_n = \{T_n = 0\}$, where $\{T_n\}_{n>0}$: tests supposed to achieve detection. Thus $\mathbb{Q}_n(E_n) \to 0$, hence by contiguity, $\mathbb{P}_n(E_n) \to 0$. Thus impossible to have $\mathbb{P}_n(E_n) \to 1$ as required by detectability.

Lemma

For $Y = \frac{d\mathbb{P}}{d\mathbb{Q}}$, $d_{var}(\mathbb{P}, \mathbb{Q}) \leq 2\sqrt{\mathbb{E}_{\mathbb{Q}} Y^2 - 1}$, *hence* $\lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}_n} Y_n^2 = 1 \Rightarrow \lim_{n\to\infty} d_{var}(\mathbb{P}_n, \mathbb{Q}_n) = 0$.

Proof:

$$\begin{aligned} d_{var}(\mathbb{P}, \mathbb{Q}) &= 2\sup_A |\mathbb{P}(A) - \mathbb{Q}(A)| = 2\sup_A |\mathbb{E}_{\mathbb{Q}}(\mathbb{I}_A(Y-1))| \\ &\leq 2\sup_A \sqrt{\mathbb{Q}(A)\mathbb{E}_{\mathbb{Q}}(Y-1)^2} \leq 2\sqrt{\mathbb{E}_{\mathbb{Q}} Y^2 - 1}. \end{aligned}$$

Detection between $\mathcal{G}(n, \alpha/n)$ and binary symmetric SBM

σ_i i.i.d. uniform on $\{-,+\}$; $\mathbb{P}((u,v) \in E | \sigma_{[n]}) = \begin{cases} \frac{a}{n} & \text{if } \sigma_u \sigma_v = +, \\ \frac{b}{n} & \text{if } \sigma_u \sigma_v = -, \end{cases}$

Spectrum of mean progeny matrix $M = \frac{1}{2}\begin{pmatrix} a & b \\ b & a \end{pmatrix}$: $\alpha = \frac{a+b}{2}$, $\beta = \frac{a-b}{2}$.

Kesten-Stigum condition: $\tau := \frac{\beta^2}{\alpha} = \frac{(a-b)^2}{2(a+b)} > 1$.

Theorem (Mossel-Neyman-Sly 2015)

Distinction between $\mathbb{P}_n : \mathcal{G}(n, \alpha/n)$ and $\mathbb{Q}_n$: symmetric binary SBM is feasible if $\tau > 1$, infeasible if $\tau < 1$.

Case $\tau > 1$: By previous theorem (spectrum of NBM B for SBM), eigenvalue of second largest modulus of $\lambda_2(B)$ verifies w.h.p.
under $\mathbb{Q}_n$: $|\lambda_2(B)| = |\beta| + o(1)$;
under $\mathbb{P}_n$, $|\lambda_2(B)| \leq \sqrt{\alpha} + o(1)$.
Hence test $T_n = \mathbb{I}_{|\lambda_2(B)|^2 \geq |\beta|\sqrt{\alpha}}$ successful at detection.

Lemma

For $Y_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$*, and* $\tau < 1$*, one has* $\mathbb{E}_{\mathbb{P}_n} Y_n^2 = (1+o(1))\frac{e^{-\tau/2-\tau^2/4}}{\sqrt{1-\tau}}$.
Thus $\sup_{n>0} \mathbb{E}_{\mathbb{P}_n} Y_n^2 < +\infty$*, hence infeasibility of detection.*

$$Y_n(g) = \sum_{s\in\{\pm\}^n} \frac{\mathbb{Q}_n(G=g;\sigma_{[n]}=s)}{\mathbb{P}_n(G=g)} = 2^{-n} \sum_{s\in\{\pm\}^n} \prod_{(uv)} W_{uv}(s), \text{ where}$$

$$W_{uv}(s) = \begin{cases} \frac{2a}{a+b} & \text{if } s_u s_v = +,\ (uv) \in E(g), \\ \frac{2b}{a+b} & \text{if } s_u s_v = -,\ (uv) \in E(g), \\ \frac{1-a/n}{1-(a+b)/2n} & \text{if } s_u s_v = +,\ (uv) \notin E(g), \\ \frac{1-b/n}{1-(a+b)/2n} & \text{if } s_u s_v = -,\ (uv) \notin E(g). \end{cases}$$

Fix $s, t \in \{\pm\}^n$. Note $W_{uv} = W_{uv}(s)$, $V_{uv} = W_{uv}(t)$, and $\epsilon = s_u s_v t_u t_v \in \{-,+\}$.

Then: $\mathbb{E}_{\mathbb{P}_n} W_{uv} V_{uv} = 1 + \epsilon\frac{(a-b)^2}{2n(a+b)} + \epsilon\frac{(a-b)^2}{4n^2} + O(n^{-3})$.

Let $\gamma := \frac{\tau}{n} + \frac{(a-b)^2}{4n^2}$, $S_\pm = |\{(uv) : s_u s v t_u t_v = \pm\}|$. Then:

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_n} Y_n^2 &= 2^{-2n} \sum_{s,t} \prod_{(uv)} \mathbb{E}_{\mathbb{P}_n} W_{uv} V_{uv} \\ &= 2^{-2n} \sum_{s,t} (1+\gamma+O(n^{-3}))^{S_+} (1-\gamma+O(n^{-3}))^{S_-}. \end{aligned}$$

Let $\rho = \rho(s,t) := \frac{1}{n} \sum_{i \in [n]} s_i t_i$.

Then: $\rho^2 = \frac{1}{n} + \frac{2}{n^2} \sum_{u \neq v} s_u s_v t_u t_v = \frac{1}{n} + \frac{2}{n^2}(S_+ - S_-)$

Also: $\frac{2}{n^2}(S_+ + S_-) = \frac{2}{n^2}\binom{n}{2} = 1 - \frac{1}{n}$. Hence:

$S_+ = (1+\rho^2)\frac{n^2}{4} - \frac{n}{2}$, $S_- = (1-\rho^2)\frac{n^2}{4}$.

For fixed $x \in \mathbb{R}$, one has: $(1+x/n)^{n^2} = (1+o(1)) e^{nx - x^2/2}$.

Thus:

$$\begin{aligned} (1+\gamma+O(n^{-3}))^{S_+} &\sim e^{-\tau/2} (1+\gamma)^{n^2(1+\rho^2)/4} \\ &\sim e^{-\tau/2} [e^{n\tau + (a-b)^2/4 - n^2\gamma^2/2}]^{(1+\rho^2)/4} \end{aligned}$$

Similarly,

$(1-\gamma+O(n^{-3}))^{S_-} \sim [e^{-n\tau - (a-b)^2/4 - n^2\gamma^2/2}]^{(1-\rho^2)/4}$

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_n} Y_n^2 &\sim 2^{-2n} e^{-\tau/2-\tau^2/4} \sum_{s,t} e^{(\rho^2/2)(n\tau+(a-b)^2/4)} \\ &\sim e^{-\tau/2-\tau^2/4} \mathbb{E} e^{(Z_n^2/2)(\tau+(a-b)^2/(4n))},\end{aligned}$$

where $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_i$, with σ_i i.i.d. uniform random signs.

By CLT, $Z_n \xrightarrow{\mathcal{W}} \mathcal{N}(0,1)$ so that $e^{Z_n^2/2(\tau+(a-b)^2/(4n))} \xrightarrow{\mathcal{W}} e^{Z^2\tau/2}$, with $Z \sim \mathcal{N}(0,1)$.

Uniform integrability of r.v. $e^{\tau Z_n^2/2}$:

$\mathbb{P}(e^{\tau Z_n^2/2} \geq M) = \mathbb{P}(|Z_n| \geq \sqrt{2\ln(M)/\tau}) \leq 2e^{-2\ln(M)/(2\tau)} = 2M^{-1/\tau}$,

by Hoeffding's inequality.

Thus

$$\begin{aligned}\mathbb{E} e^{\tau Z_n^2/2} \mathbb{I}_{e^{\tau Z_n^2/2} \geq M} &= M\mathbb{P}(e^{\tau Z_n^2/2} \geq M) + \int_M^\infty \mathbb{P}(e^{\tau Z_n^2/2} \geq x)dx \\ &\leq 2M^{1-1/\tau} + 2\int_M^\infty x^{-1/\tau} dx \\ &\to 0 \text{ as } M \to \infty \text{ for } \tau < 1.\end{aligned}$$

By uniform integrability, $\mathbb{E}_{\mathbb{P}_n} Y_n^2 \to \mathbb{E} e^{Z^2\tau/2} = \frac{1}{\sqrt{1-\tau}}$.

The planted clique detection problem

Under H_0 ($\mathbb{P}_n$), Erdős-Rényi graph $G \sim \mathcal{G}(n, 1/2)$;
Under H_1 ($\mathbb{Q}_n$), $G : \mathcal{G}(n, 1/2) \cup$ clique of size K, on subset $\mathcal{K}$ uniformly chosen in $\binom{[n]}{K}$.

Theorem (Informational threshold)

For any $\epsilon > 0$, if $K \leq (1-\epsilon) 2 \log_2(n)$ then $d_{var}(\mathbb{P}_n, \mathbb{Q}_n) \to 0$, hence detection infeasible.
If $K \geq K_0 := (1+\epsilon) 2 \log_2(n)$, detection feasible, based on test $T = \mathbb{I}_{G \text{ contains clique of size } K_0}$.

Proof (infeasibility):

$$\begin{aligned} Y_n(g) &= \frac{\mathbb{Q}_n(G=g)}{\mathbb{P}_n(G=g)} = \frac{1}{\binom{n}{K}} \sum_{C \in \binom{[n]}{K}} \frac{\mathbb{Q}_n(G=g|\mathcal{K}=C)}{\mathbb{P}_n(G=g)} \\ &= \frac{1}{\binom{n}{K}} \sum_{C \in \binom{[n]}{K}} 2^{\binom{K}{2}} \mathbb{I}_{C \text{ clique of } g} \end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_n} Y_n^2 &= \frac{1}{\binom{n}{K}^2} \sum_{C,C' \in \binom{[n]}{K}} 2^{2\binom{K}{2}} \mathbb{P}_n(C, C' \text{ cliques of } G) \\
&= \frac{1}{\binom{n}{K}} \sum_{C \in \binom{[n]}{K}} 2^{2\binom{K}{2}} \mathbb{P}_n(C, [K] \text{ cliques of } G) \\
&= \frac{1}{\binom{n}{K}} \sum_{k=0}^{K} 2^{2\binom{K}{2}} \binom{K}{k} \binom{n-K}{K-k} \left(\frac{1}{2}\right)^{2\binom{K}{2} - \binom{k}{2}} \\
&\leq \frac{1}{\binom{n}{K}} \sum_{k=0}^{K} \binom{K}{k} \binom{n-K}{K-k} 2^{kK/2}
\end{aligned}$$

Since $K \leq (1-\epsilon)2\log_2(n)$, $2^{kK/2} \leq n^{(1-\epsilon)k}$, so that

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_n} Y_n^2 &\leq (1+o(1)) \frac{K!}{n^K} \sum_{k=0}^{K} \binom{K}{k} \frac{n^{K-k}}{(K-k)!} n^{(1-\epsilon)k} \\
& (1+o(1)) \sum_{k=0}^{K} \binom{K}{k} \frac{K!}{(K-k)!} n^{-\epsilon k}
\end{aligned}$$

Let $f(k) = \binom{K}{k} \frac{K!}{(K-k)!} n^{-\epsilon k}$. Then $\frac{f(k+1)}{f(k)} = \frac{(K-k)^2}{k+1} n^{-\epsilon} < 1$ for large enough n.

Hence:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_n} Y_n^2 &\leq (1+o(1))[1 + Kf(1)] \\
&= (1+o(1))[1 + K^3 n^{-\epsilon}] \\
&= 1 + o(1).
\end{aligned}$$

Proof (feasibility):

For $K \geq k_0 := (1+\epsilon)2\log_2(n)$, and $T_n = \mathbb{I}_{G \text{ contains a } k_0\text{-clique}}$, obviously $\mathbb{Q}_n(T_n = 1) = 1$.

Write

$$
\begin{aligned}
\mathbb{P}_n(T_n = 1) &\leq \binom{n}{k_0} 2^{-\binom{k_0}{2}} \\
&\leq n^{k_0} 2^{-k_0(k_0-1)/2} \\
&\leq n^{k_0} n^{-(1+\epsilon)(k_0-1)} \\
&\leq n^{-\epsilon k_0 + 1 + \epsilon} \\
&\to 0 \text{ as } n \to \infty.
\end{aligned}
$$

Remark

No known polynomial-time implementation of this test: finding a maximum size-clique in a graph is NP-hard (it is even NP-hard to find an approximation of the size of the largest clique up to an approximation factor $n^{1-\epsilon}$ *for any* $\epsilon \in (0,1)$, *[Hastad, 1999]).*

Planted clique detection, computational threshold

Theorem (Alon-Krivelevitch-Sudakov, 1998)

For constant $c > 0$, detection of planted cliques of size $K = c\sqrt{n}$ can be done in polynomial time, using a spectral method.

To graph G with adjacency matrix A, associate matrix $W_{uv} = 2A_{uv} - 1$.
Let $G^0 \sim \mathcal{G}(n, 1/2)$, with adjacency matrix A^0 and for $u \neq v \in [n]$ let $W^0_{uv} = 2A^0_{uv} - 1$.
For set $\mathcal{K} \subset [n]$, $|\mathcal{K}| = K = c\sqrt{n}$,
let $\Delta_{uv} = \begin{cases} 0 & \text{if } u \text{ or } v \notin \mathcal{K}, \\ 1 - W^0_{uv} & \text{if both } u, v \in \mathcal{K}. \end{cases}$
Then for $G^1 = G^0 \cup$ clique on $\mathcal{K}$, $W^1 = W^0 + \Delta$.
Note also: $\bar{\Delta}_{uv} = \mathbb{I}_{u,v \in \mathcal{K}}$.

Theorem (Anderson-Guionnet-Zeitouni, proof of Theorem 2.1.22)

For $(Y_i)_{i\in[n]}$ *i.i.d.,* $(Z_{ij})_{i<j\in[n]}$ *i.i.d.,* $\mathbb{E}Y_1 = \mathbb{E}Z_{12} = 0$, $\mathbb{E}Z_{12}^2 = 1$, *let Wigner matrix* $W^{(n)}$*:* $W_{ii}^{(n)} = n^{-1/2}Y_i$, $W_{ij}^{(n)} = W_{ji}^{(n)} = n^{-1/2}Z_{ij}$, $i < j \in [n]$.

Assume $r_k := \max(\mathbb{E}|Z_{12}|^k, \mathbb{E}|Y_1|^k) \le k^{ak}$ *for some constant* $a > 0$*. Then for all* $\delta > 0$*:*

$\mathbb{P}(\rho(W^{(n)}) \le 2 - \delta) \overset{n\to\infty}{\to} 0$,

and for any constant $b > 0$,

$\mathbb{P}(\rho(W^{(n)}) \ge 2 + \delta) = o(n^{-b})$.

Thus w.h.p., $\rho(W^0) = (1 + o(1))2\sqrt{n}$ (W^0: Wigner matrix of size n), and $\rho(\Delta - \bar{\Delta}) = (1 + o(1))2\sqrt{K}$ (conditionally on $\mathcal{K}$, $\Delta - \bar{\Delta}$: Wigner block of size K)

$\bar{\Delta}$: rank 1 matrix with spectral radius K, and associated eigenvector $x = (\mathbb{I}_{u\in\mathcal{K}})_{u\in[n]}$.

Write $W^1 = W^0 + (\Delta - \bar{\Delta}) + \bar{\Delta} = \tilde{W} + \bar{\Delta}$, where $\tilde{W} = W^0 + (\Delta - \bar{\Delta})$.
Thus w.h.p., $\rho(\tilde{W}) \leq (1 + o(1))2[\sqrt{n} + \sqrt{K}] = (1 + o(1))2\sqrt{n}$
Assume $K = c\sqrt{n}$, $c > 4$. Then:
Under $\mathbb{P}_n$, $\rho(W^0) = (1 + o(1))2\sqrt{n}$, and by Weyl's inequalities, under $\mathbb{Q}_n$, $\rho(W^1) \geq K - (1 + o(1))2\sqrt{n} \geq (c - 2)(1 + o(1))\sqrt{n}$.
Hence test $T_n = \mathbb{I}_{\rho(W) > (1+\epsilon)2\sqrt{n}}$ detects w.h.p. between H_0 and H_1.

For $c > 0$ not necessarily > 4:
For each set of size ℓ of nodes $\{i_1, \ldots, i_\ell\} \in [n]$, consider subgraph $G^{i_1^\ell}$ of G among nodes i neighbours of all $i_1, \ldots, i_\ell$.
Then under H_0, for each choice $i_1, \ldots, i_\ell$, with probability $1 - o(n^{-\ell})$, $\rho(W^{i_1^\ell}) \leq (1 + o(1))2\sqrt{2^{-\ell} n}$
Thus (union bound) w.h.p., for all $i_1^\ell \in [n]$,$\rho(W^{i_1^\ell}) \leq (1 + o(1))2\sqrt{2^{-\ell} n}$.
Under H_1, for choice $G^{i_1^\ell}$: $\sim 2^{-\ell} n$ vertices, planted clique of size $c\sqrt{n}$.
Hence $\rho(W^{i_1^\ell}) \geq [c - 2(1 + o(1))2^{-\ell/2}]\sqrt{n}$.
Thus for $\ell : c > 4\sqrt{2^{-\ell}}$, test $T_n = \mathbb{I}_{\exists i_1^\ell \in [n] : \rho(W^{i_1^\ell}) \geq (1+\epsilon)2\sqrt{2^{-\ell} n}}$ detects between H_0 and H_1.

Remark

No polynomial-time algorithm known for detection of planted clique of size $K = o(\sqrt{n})$. *Conjecture: no polynomial-time algorithm exists in that case. Hence three phases: detection is*

- *infeasible for* $K \leq (1-\epsilon) 2 \log_2(n)$,
- *feasible (in an information-theoretic sense), but computationally hard for* $(1+\epsilon) 2 \log_2(n) \leq K \leq o(\sqrt{n})$,
- *poly-time feasible for* $K = \Omega(\sqrt{n})$.

Clique reconstruction under H_1 and $K = c\sqrt{n}$, $c > 4$

$\rho(\bar{\Delta}) = c\sqrt{n}$, and $\rho(\bar{\Delta})$ separated from other eigenvalue 0 by $c\sqrt{n}$.
$\rho(\tilde{W}) \leq 2(1 + o(1))\sqrt{n}$.

By results on perturbation of eigenvectors, $\exists x \leftrightarrow \rho(W^1)$ such that
$\langle x, \bar{x} \rangle \geq \sqrt{1 - \frac{\rho(\tilde{W})^2}{(c\sqrt{n} - \rho(\tilde{W}))^2}} = \Omega(1)$, where $\bar{x}_u = \frac{1}{\sqrt{K}} \mathbb{I}_{u \in \mathcal{K}}$
Hence $\sum_u x_u^2 = 1$, $\frac{1}{\sqrt{K}} \sum_{u \in \mathcal{K}} x_u \geq \beta = \sqrt{1 - [2/(c-2)]^2} + o(1) = \Omega(1)$.

Let for constant $a \in (0, \beta)$: $C = \{u \in \mathcal{K} : x_u \geq \frac{a}{\sqrt{K}}\}$. Thus:
$\sqrt{K}\beta \leq \sum_{u \in \mathcal{K}} x_u \leq (K - |C|)\frac{a}{\sqrt{K}} + \sum_{u \in C} x_u \leq K\frac{a}{\sqrt{K}} + \sqrt{|C|}$

Hence $\sqrt{K}(\beta - a) \leq \sqrt{|C|}$, i.e. $|C| \geq K(\beta - a)^2$.

Let $D = \{u \in \bar{\mathcal{K}} : x_u \geq \frac{a}{\sqrt{K}}\}$. Necessarily, $|C| + |D| \leq \frac{K}{a^2}$.
Thus among nodes in $E = \{u : x_u \geq a/\sqrt{K}\} = C \cup D$, fraction at least $a^2(\beta - a)^2 = \Omega(1)$ belongs to $\mathcal{K}$.

Set aside set $\tilde{V}$ of ϵn vertices chosen at random, $\epsilon > 0$. Remaining graph G': $n(1-\epsilon)$ vertices, and planted clique of size $c\sqrt{n}(1-\epsilon-o(1))$ whp.

Previous analysis applies to G' for ϵ such that $c\sqrt{1-\epsilon} > 4$: set E' of size $m = \Theta(K)$ contains fraction $\alpha = \Omega(1)$ of clique vertices in G'.

For $u \in \tilde{V} \cap \mathcal{K}$, $X_u := \sum_{v \in E'} \mathbb{I}_{u \sim v} \geq \alpha m + \text{Bin}((1-\alpha)m, 1/2)$;
For $u \in \tilde{V} \setminus \mathcal{K}$, $X_u := \sum_{v \in E'} \mathbb{I}_{u \sim v} = \text{Bin}(m, 1/2)$;
Let $\hat{\mathcal{K}} := \{u \in \tilde{V} : X_u \geq (1-\alpha)\frac{m}{2} + \frac{2}{3}\alpha m\}$.

Then Chernoff bounds for binomial random variables imply:
whp, $\hat{\mathcal{K}} = \tilde{V} \cap \mathcal{K}$.
$\rightarrow$ whp, exact reconstruction of $\mathcal{K}$ on $\tilde{V}$.

Reconstruction of $\mathcal{K}$ on $V' = [n] \setminus \tilde{V}$: whp, nodes in V' neighbours of all nodes in $\hat{\mathcal{K}}$ are exactly nodes in $V' \cap \mathcal{K}$.

Remark

Reconstruction in polynomial time also feasible for $K = c\sqrt{n}$, $c < 4$*: identify subgraph* $G^{i_1^\ell}$ *with* $2^{-\ell}n$ *vertices and planted clique of size* $c\sqrt{n}$*, where* $c > 4.2^{-\ell/2}$*. Use previous reconstruction algorithm of* $\hat{\mathcal{K}}$ *on this graph, giving whp subset of size* $\Theta(\sqrt{n})$ *of* $\mathcal{K}$*. Add to it all nodes that are neighbours of everyone in* $\hat{\mathcal{K}}$*.*