For a given graph property, e.g., bipartiteness, and graph $G$ want to distinguish between the cases:

1. $G$ has the property.
2. $G$ is “$\epsilon$-far” from having the property, i.e., need to change $\geq \epsilon n^2$ entries of adjacency matrix to get graph with the property.

For some problems, a more stringent definition of $\epsilon$-far will be used.

Our focus will be on the number of queries we need to make to the graph where a query is of the form “is $(u, v)$ an edge?”
Bipartiteness

For a given graph $G$, want to distinguish the cases:

1. There exists a partition of the nodes $V_1, V_2$ with no violating edges, i.e., edges with both end points in the same $V_i$
2. All partitions of the nodes have at least $\epsilon n^2$ violating edges.

Basic Idea:

- Sample $r = \epsilon^{-1} \ln(1/\delta)$ pairs of nodes $(u_1, v_1), (u_2, v_2), \ldots$.
- For a given partition with $\geq \epsilon n^2$ violating edges, the probability one of them is violating for this partition is at least:
  \[ 1 - (1 - \epsilon)^r \geq 1 - e^{-\epsilon r} = 1 - \delta \]

Can turn this into an algorithm:

- Given the $r$ samples, try every node partition and if there exists a partition with no violating edge, accept and otherwise, reject.
- Probability of error is at most $2^n \delta$ so setting $\delta = 1/(4 \cdot 2^n)$ ensure error probability $1/4$ with $r = \epsilon^{-1}(n + 2) \ln 2$ samples.
A more efficient Bipartiteness tester

Let’s not try all partitions, but only a random set of partitions:

1. Let $U$ be a random set of $\Theta(\epsilon^{-1} \log \epsilon^{-1})$ nodes.
   1.1 If $U$ is not bipartite output “reject”

2. For each partition $U_1, U_2$ of $U$, consider induced partition of $V$

   \[ W_1 = U_1 \cup \Gamma(U_2) \cup \{ v \in V \setminus U : v \notin \Gamma(U_1) \cup \Gamma(U_2) \} \]
   \[ W_2 = U_2 \cup \Gamma(U_1) \]

2.1 If we observe $v \in \Gamma(U_1) \cap \Gamma(U_2)$, we can ignore $(U_1, U_2)$.

Note that we don’t need to compute a partition explicitly: given $U_1, U_2$, we can determine whether $v \in W_1$ or $v \in W_2$ using $|U|$ queries.

Can show if $G$ is bipartite there’s a partition defined above with $\leq \epsilon n^2/2$ violated edges. If not, all partitions have $\geq \epsilon n^2$ violated edges. Can distinguish with

\[
r = \Theta(\epsilon^{-1} \log \delta^{-1})
\]

queries via Chernoff analysis and now we only need $\delta = O(1/2|U|)$. 

4/12
If $G$ is bipartite, let $(Y_1, Y_2)$ be a partition with no violating edges.

For sampled nodes $U$, consider partition:

\[ U_1 = U \cap Y_1 \quad U_2 = U \cap Y_2 \]

and associated partition $(W_1, W_2)$.

How different are $(Y_1, Y_2)$ and $(W_1, W_2)$? Note that all nodes on wrong side must have had no neighbors in $U$. 
If $G$ is bipartite, there's a partition with few violations (2/2)

- Suppose there's $A$ wrong-side nodes with degree $<$ $\epsilon n/4$ and $B$ wrong-side with degree $\geq \epsilon n/4$.
- Number of violating edges in $(W_1, W_2)$ is
  \[
  |A|\epsilon n/4 + |B|n \leq \epsilon n^2/4 + |B|n
  \]
- If $|B| \leq \epsilon n/4$, number of violating edges in $(W_1, W_2)$ is $\leq \epsilon n^2/2$.
- Probability node with degree $\geq \epsilon n/4$ has no neighbors in $U$:
  \[
  (1 - (\epsilon n/4)/n)^{|U|} \leq \exp(-\epsilon/4 \cdot |U|) \leq \epsilon/12
  \]
- $\mathbb{E}[|B|] \leq \epsilon n/12$ and by Markov bound, with probability at least $2/3$, $|B| \leq \epsilon n/4$
Definition: $H$ is a minor of a graph $G$ if we can obtain $H$ via a sequence of vertex removals, edge removals, or edge contractions. $G$ is $H$-minor free if $H$ is not a minor of $G$.

Wagner’s Thm: $G$ is planar iff it is $K_{3,3}$ and $K_5$ minor free.

A graph property $P$ is minor-closed if $G$ having the property implies all minors of $G$ also have the property. Can be shown every minor closed property is expressible as a constant number of excluded minors.

We’ll develop an algorithm running in $f(\epsilon, d)$ time that for any $G$ and constant size $H$, distinguishes between the cases:

- $G$ is $H$-minor-free.
- $G$ is $\epsilon$-far from being $H$-minor-free, i.e., more than $\epsilon dn$ edges need removed where $d$ is max degree of $G$. 

(\(\epsilon, k\))-Hyperfinite

- \(G\) is \((\epsilon, k)\)-hyperfinite if it is possible to remove \(\leq \epsilon n\) edges such that the resulting connected components all have at most \(k\) nodes.

- **Theorem:** Given \(H\), there exists constant \(C_H\) such that \(\forall \epsilon \in (0, 1)\), every \(H\)-minor free graph of degree \(\leq d\) is \((\epsilon d, C_H^2/\epsilon^2)\)-hyperfinite.

- Using hyperfiniteness:
  - Partition \(G\) into a graph \(G'\) with constant size components by removing few edges. If we can’t do this, \(G\) is not \(H\)-minor free.
  - Test if \(G'\) has the property by picking random components and checking if they have the property.

- Initially assume we have a partition oracle:
  - **Input:** vertex \(v\).
  - **Output:** \(P[v]\), i.e., name of partition including \(v\).
  - **Properties of partition:**
    - 1) For \(v\), \(\{u : P[u] = P[v]\} \leq k\)
    - 2) if \(G\) is \(H\)-minor free
      \[
      |\{(u, v) \in E : P[u] \neq P[v]\}| \leq \epsilon dn/4
      \]
Algorithm assuming partition oracle

1. Part I: Does partition oracle only have a few crossing edges
   ▶ Compute estimate $\hat{f}$ of $|\{(u, v) \in E : P[u] \neq P[v]\}|$. Estimate should be correct up to additive error $\epsilon dn/8$ with probability $\geq 9/10$
   ▶ If $\hat{f} > 3/8 \cdot \epsilon dn$ conclude $G$ is not $H$-minor-free

2. Part II: Test random partitions
   ▶ Chose $S = O(1/\epsilon)$ random nodes and construct components corresponding to partitions including nodes in $S$.
   ▶ If all components explored are $H$-minor free and have size $\leq k = O(1/\epsilon^2)$, conclude $G$ is close to $H$-minor-free.

Runtime analysis:
   ▶ Part I requires $O(1/\epsilon^2)$ calls to oracle.
   ▶ Part II requires $O(d/\epsilon^2)$ calls to oracle to construct each component and so $O(d/\epsilon^3)$ in total.
Correctness Analysis

- If $G$ is $H$-minor free:
  - $|\{(u, v) \in E : P[u] \neq P[v]\}| \leq \epsilon dn/4$ implies
    \[
    \hat{f} \leq \epsilon dn/4 + \epsilon dn/8 = 3d\epsilon n/8
    \]
    with probability at least $9/10$. So we probably pass Part I.
  - For all $u \in V$, $P[u]$ is small and $H$ minor free. So we pass Part II.

- If $G$ is $\epsilon$-far from $H$-minor free:
  - Case 1: If $|\{(u, v) \in E : P[u] \neq P[v]\}| > \epsilon dn/2$ then
    \[
    \hat{f} > \epsilon dn/2 - \epsilon dn/8 = 3d\epsilon n/8
    \]
    with probability at least $9/10$. So we'd probably reject at Part I.
  - Case 2: If $|\{(u, v) \in E : P[u] \neq P[v]\}| \leq \epsilon dn/2$.
    - We know $G'$ is $\epsilon/2$-far from having the property.
    - The $\geq \epsilon dn/2$ edges needing removed are incident to $\geq \epsilon n/2$ nodes.
    - Sampling $O(1/\epsilon)$ nodes suffices to find component with an $H$ minor.
How do we implement the partition oracle? A sketch.

Definition: $S$ is an $(\delta, k)$-isolated neighborhood of a node $v$ if

1. $v \in S$
2. $S$ is connected.
3. $|S| \leq k$
4. number of edges between $S$ and $\bar{S}$ is at most $\delta |S|$

At least $(1 - \epsilon/30)n$ nodes of a $H$-minor-free graph $G$ have a $(\delta, k)$ isolated neighborhood for $\delta = d\epsilon/30$ and $k = \Theta(1/\epsilon^4)$

1. Since $G$ is $H$-minor-free, it is possible to remove $r = 2\epsilon^2dn/900$ edges leaving connected components is size $\leq k = O(1/\epsilon^4)$.
2. Say a node is bad if its partition is not a $(\delta, k)$-isolated neighborhood. Then number of bad nodes $n_b$ satisfies:

$$\delta n_b \leq r/2 = \epsilon^2dn/900$$

and so $n_b \leq \epsilon/30 \cdot n$ as required.
Consider the (not sublinear time) greedy algorithm:

- Process nodes in random order and for each $v$ explore locally to find a $(\delta, k)$ isolated neighborhood. If we find one, define this to be partition for $v$. If not, let partition of $v$ just be $\{v\}$.
- By previous slide, expected number of singleton partitions is at most $\epsilon n/30$ and is less than $\epsilon n/10$ with probability at least $2/3$.
- Number of edges removed to define this partition is at most

$$\delta n + d \cdot (\#\text{singleton partitions}) \leq \epsilon d n/30 + d \epsilon n/10 \leq 2\epsilon d n/15$$

- Can emulate this greedy algorithm using similar ideas to those used to construct greedy matching.