Extremal Graph Theory

- The subfield of **extremal graph theory** deals with questions of the form:
  - What is the **maximum number of edges** that a graph with $n$ vertices can have **without containing a given subgraph $H$**?
  - Characterize the graphs that obtain the this maximum number of edges.
**ex(n, H)**

- We denote by \( ex(n, H) \) the maximum number of edges in a graph with \( n \) vertices and no subgraph \( H \).

- **Example.** Recall that \( P_i \) is a simple path of length \( i \). What is \( ex(4, P_3) \)?

\[ \begin{aligned}
\text{\small \begin{tikzpicture}
  \begin{scope}
    \node (v1) at (0,0) [circle,fill=blue] {};
    \node (v2) at (0.5,0.5) [circle,fill=blue] {};
    \node (v3) at (1,0) [circle,fill=blue] {};
    \draw (v1) -- (v2);
  \end{scope}
  \begin{scope}[xshift=1cm]
    \node (v1) at (0,0) [circle,fill=blue] {};
    \node (v2) at (0.5,0) [circle,fill=blue] {};
    \node (v3) at (1,0) [circle,fill=blue] {};
    \draw (v1) -- (v2);
  \end{scope}
\end{tikzpicture}} \end{aligned} \]

**Graphs Without Triangles**

- **Problem.** Find \( ex(n, K_3) \).
  - What graph contains a large number of edges but does not have \( K_3 \) as a subgraph?
  - Any complete bipartite graph.

\[ \begin{aligned}
\text{\small \begin{tikzpicture}
  \begin{scope}
    \node (v1) at (0,0) [circle,fill=red] {};
    \node (v2) at (0,1) [circle,fill=red] {};
    \node (v3) at (0,2) [circle,fill=red] {};
    \draw (v1) -- (v2) -- (v3);
  \end{scope}
  \begin{scope}[xshift=3cm]
    \node (v1) at (0,0) [circle,fill=red] {};
    \node (v2) at (0,1) [circle,fill=red] {};
    \node (v3) at (0,2) [circle,fill=red] {};
    \draw (v1) -- (v2) -- (v3);
  \end{scope}
\end{tikzpicture}} \end{aligned} \]
Maximizing the Number of Edges

• What value of \( m \) maximizes the number of edges in \( K_{m,n-m} \)?
  ◦ The number of edges is \( m(n-m) \).
  ◦ The maximum of \( n^2/4 \) is obtained for \( G_{n/2,n/2} \).
  ◦ Can we find a graph without triangles and with more edges?

Mantel’s Theorem

• **Theorem.** \( \text{ex}(n,K_3) = [n^2/4] \).
• **Proof.** Let \( G = (V,E) \) be a triangle-free graph with \( |V| = n \).
  ◦ \( d_i = \deg v_i \).
  ◦ If \((v_i,v_j) \in E\), then \( v_i \) and \( v_j \) do not have common neighbors, so \( d_i + d_j \leq n \).
  ◦ We thus have \( \sum_{(v_i,v_j) \in E} (d_i + d_j) \leq n|E| \).
  ◦ Since every \( d_i \) appears in \( d_i \) elements
    \[ n|E| \geq \sum_{(d_i,d_j) \in E} (d_i + d_j) = \sum_i d_i^2 . \]
The Cauchy–Schwarz Inequality

- **The Cauchy–Schwarz inequality.** For any \( a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{R} \), we have
  \[
  \left( \sum_{i=1}^{n} a_i b_i \right)^2 \leq \left( \sum_{i=1}^{n} a_i^2 \right) \left( \sum_{i=1}^{n} b_i^2 \right).
  \]

- Equality holds iff the vectors \((a_1, \ldots, a_n)\) and \((b_1, \ldots, b_n)\) are linearly dependent.

Available online

Completing the Proof

- We proved that
  \[
  n|E| \geq \sum_{(d_i, d_j) \in E} (d_i + d_j) = \sum_i d_i^2
  \]
- Recall that \( \sum_i d_i = 2|E| \).
- **By Cauchy–Schwarz** with \( a_i = d_i \) and \( b_i = 1 \):
  \[
  \left( \sum_i d_i \right)^2 \leq \left( \sum_i d_i^2 \right) \left( \sum_i 1^2 \right).
  \]
  - We thus have
  \[
  n|E| \geq \sum_i d_i^2 \geq \frac{4|E|^2}{n} \quad \rightarrow \quad |E| \leq \frac{n^2}{4}
  \]
The Maximum Graph is Unique

- **Claim.** The only graph that has the maximum number of edges \( ex(n, K_3) = \lfloor n^2 / 4 \rfloor \) is \( K_{[n/2],[n/2]} \).

- **Proof.** We consider the case of an even \( n \).

  - In our proof for \( ex(n, K_3) = n^2 / 4 \), we used **Cauchy-Schwarz** to obtain 
    \[ (\sum_i d_i)^2 \leq (\sum_i d_i^2)(\sum_i 1^2), \]
    which implied 
    \[ n|E| \geq \sum_i d_i^2 \geq \frac{4|E|^2}{n}. \]

  - Equality holds if and only if for every \( 1 \leq i \leq n \), we have \( d_i = n/2 \).

Inequality of Arithmetic and Geometric Means

- **AM-GM Inequality.** For any \( a_1, a_2, \ldots, a_n \) \( \in \mathbb{R} \), we have 
  \[ \frac{a_1 + \cdots + a_n}{n} \geq (a_1 \cdots a_n)^{1/n}. \]
Recall: Independent Sets

- Consider a graph $G = (V, E)$. An \textbf{independent set} in $G$ is a subset $V' \subset V$ such that there is no edge between any two vertices of $V'$.

- Finding a \textbf{maximum independent set} in a graph is a major problem in theoretical computer science.
  - No polynomial-time algorithm is known.

Mantel’s Theorem: Second Proof

- $G = (V, E)$ – a \textbf{triangle-free} graph with $|V| = n$.
- $\alpha$ – the size of the largest independent set $A$ in $G$.
- The set of neighbors of any $v \in V$ is an \textbf{independent set}, so $\alpha \geq \deg v$.
- $B = V \setminus A$, so $|B| = n - \alpha$. Also, $B$ is a \textbf{vertex cover}.
- \textbf{By the AM-GM inequality}, we have
  \[
  |E| \leq \sum_{v \in B} \deg v \leq \alpha |B| \leq \left(\frac{\alpha + (n - \alpha)}{2}\right)^2 = \frac{n^2}{4}.
  \]
Generalizing the Problem

- We move from $\text{ex}(n, K_3)$ to $\text{ex}(n, K_r)$.
  - For some $r > 3$, what graph contains many edges but no $K_r$?
  - An $r$-partite graph is a graph with $r$ parts, and no edge between two vertices of the same part (generalizing bipartite graphs).
  - \textbf{Example.} The figure presents the complete 4-partite graph $K_{3,3,3,4}$.

Turán Graphs

- The \textbf{Turán graph} $T(n, r)$ is a complete $r$-partite graph with $n$ vertices, such that each part consists of either $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$ vertices.
  - The graph in the figure is $T(13,4)$.
  - $T(n, r - 1)$ contains no copy of $K_r$ and has about \( n \left( n - \frac{n}{r-1} \right) / 2 \) edges.
Turán’s Theorem

• **Theorem.** If \( G = (V, E) \) is a graph that contains no copy of \( K_r \) and \( |V| = n \), then \( |E| \leq \frac{n^2}{2} \left(1 - \frac{1}{r-1}\right) \).

• **Proof.** By induction on \( n \).
  - **Induction basis.** Easily holds when \( n < r \).
  - **Induction step.** \( G = (V, E) \) – a graph that contains no copy of \( K_r \), with \( |V| = n \) and \( |E| = ex(n, K_r) \).
  - \( G \) contains a copy of \( K_{r-1} \). Otherwise, adding an edge to \( G \) would not yield a copy of \( K_r \), **contradicting the maximality of \( G \)**.

• \( G = (V, E) \) – a graph that contains no copy of \( K_r \), with \( |V| = n \) and \( |E| = ex(n, K_r) \) edges.
• \( A \) – an induced subgraph of \( G \) that is a \( K_{r-1} \).
• \( A \) contains \( \binom{r-1}{2} \) edges.

  - **By the hypothesis,** \( G \setminus A \) has at most \( \frac{(n-r+1)^2}{2} \left(1 - \frac{1}{r-1}\right) \) edges.
  - Since each vertex of \( V \setminus A \) is adjacent to at most \( r-2 \) vertices of \( A \), there are \( (n-r+1)(r-2) \) edges between \( V \setminus A \) and \( A \).
  - Combining the above, we have
    \[
    |E| \leq \frac{(r-1)(r-2)}{2} + \frac{(n-r+1)^2}{2} \left(1 - 1/(r-1)\right) + (n-r+1)(r-2) \leq \frac{n^2}{2} \left(1 - 1/(r-1)\right).
    \]
Which Actor from the Big Bang Theory is a Real Scientist?

Proof #2: Balancing Weights

- \( G = (V, E) \) – a graph that contains no copy of \( K_r \), with \( V = \{v_1, ..., v_n\} \).
- We assign a weight \( w_i = \frac{1}{n} \) to every \( v_i \). We set \( W = \sum_{(v_i, v_j) \in E} w_i w_j \).
- Initially we have \( W = |E| \frac{1}{n^2} \). We wish to move weights to increase \( W \).
Balancing Weights (cont.)

- \(v_i, v_j\) – two vertices of \(V\) with positive weights, such that \((v_i, v_j) \notin E\).
- \(N_i, N_j\) – the sum of the weights of the neighbors of \(v_i\) and \(v_j\), respectively.
- WLOG, assume that \(N_i \leq N_j\).
- By moving the weight of \(v_i\) to \(v_j\), we change \(W\) by \(w_i N_j - w_i N_i \geq 0\).

More Balancing of Weights.

- If there exist \(v_i, v_j\) with positive weights and \((v_i, v_j) \notin E\), we can move the weight vertex of one to the other without decreasing \(W\).
- We repeatedly do this until the entire weight is on the vertices of a complete subgraph \(K_m\).
- Consider \(v_i, v_j \in K_m\) with \(w_j - w_i \geq 2\varepsilon > 0\). Then moving \(\varepsilon\) from \(w_j\) to \(w_i\) changes \(W\) by
  \[\varepsilon \left( (1 - w_i - \varepsilon) - (1 - w_j) \right) = \varepsilon (w_j - w_i - \varepsilon) \geq \varepsilon^2.\]
Concluding the Proof

- We conclude that we can move the entire weight to a subgraph $K_m$ where each vertex of the subgraph has a weight of $\frac{1}{m}$.

  - By the assumption, $m \leq r - 1$.

  - That is, we obtain $W \leq \binom{r - 1}{2} \frac{1}{(r-1)^2} = \frac{r-2}{2r-2}$.

  - Recall that we started with $W = \frac{1}{n^2}$.

  - Since we only increased $W$, we have
    \[
    |E| \frac{1}{n^2} \leq \frac{r-2}{2r-2} \quad \Rightarrow \quad |E| \leq \frac{n^2}{2} \left(1 - \frac{1}{r-1}\right).
    \]

Cliques

- Given a graph $G = (V, E)$, a **clique** of $G$ is any subgraph that is a complete graph.

- No polynomial-time algorithm for finding a **maximum clique** in a graph is known.
  \- The problem is equivalent to finding a **maximum independent set** in a graph.
  \\- Define the graph $G^c = (V, E^c)$ such that $(v, u) \in E^c$ iff $(v, u) \notin E$. Finding a maximum clique in $G$ is equivalent to finding a maximum independent set in $G^c$. 
A Lower Bound on the Clique Size

**Lemma.** Let $G = (V, E)$ be a graph with $V = \{v_1, \ldots, v_n\}$, and let $d_i = \deg v_i$. Then $G$ contains a clique of size at least $\sum_{i=1}^{n} \frac{1}{n-d_i}$.

\[
\frac{1}{5-4} + \frac{1}{5-4} + \frac{1}{5-3} + \frac{1}{5-3} + \frac{1}{5-2} = \frac{3}{3} + \frac{1}{3}
\]

Probabilistic Proof

- We **uniformly** choose a permutation $\pi$ of $\{1, \ldots, n\}$.
- We build a subset $C_\pi \subset V$: for every $1 \leq i \leq n$, we add $v_{\pi(i)}$ to $C_\pi$ if $v_{\pi(i)}$ is adjacent to $v_{\pi(1)}, v_{\pi(2)}, \ldots, v_{\pi(i-1)}$.
- Notice that $C_\pi$ is a **clique**.
- $X_i$ – an **indicator variable** for whether $v_{\pi(i)} \in C_\pi$.
- $E[X_i] = \Pr[X_i = 1] = \frac{1}{n-d_{\pi(i)}}$. 
Proof (cont.)

- \( X_i \) – an indicator variable for whether \( v_{\pi(i)} \in C_\pi \).
- \( E[X_i] = Pr[X_i = 1] = \frac{1}{n - d_{\pi(i)}} \).
- Set \( X = \sum_{i=1}^{n} X_i \).
- \( X \) is the size of the clique \( C_\pi \).
- By linearity of expectation

\[
E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{n - d_{\pi(i)}} = \sum_{i=1}^{n} \frac{1}{n - d_i}.
\]

Completing the Proof

- \( X \) is the size of the clique \( C_\pi \).

\[
E[X] = \sum_{i=1}^{n} \frac{1}{n - d_i}.
\]

- Thus, there exists a permutation that leads to a clique of at least this size.
Proof #3 of Turán’s Theorem

- \( G = (V, E) \) – a graph that contains no copy of \( K_r \), with \( V = \{v_1, ..., v_n\} \).

\[ d_i = \text{deg} v_i, \quad a_i = \sqrt{n - d_i}, \quad b_i = \frac{1}{\sqrt{n - d_i}}. \]

- By Cauchy-Schwarz

\[ \left( \sum_{i=1}^{n} a_i b_i \right)^2 \leq \left( \sum_{i=1}^{n} a_i^2 \right) \left( \sum_{i=1}^{n} b_i^2 \right). \]

- Since \( a_i b_i = 1 \), we have

\[ n^2 \leq \left( \sum_{i=1}^{n} a_i^2 \right) \left( \sum_{i=1}^{n} b_i^2 \right). \]

\[ d_i = \text{deg} v_i, \quad a_i = \sqrt{n - d_i}, \quad b_i = \frac{1}{\sqrt{n - d_i}}. \]

- We have

\[ n^2 \leq \left( \sum_{i=1}^{n} a_i^2 \right) \left( \sum_{i=1}^{n} b_i^2 \right) = \left( \sum_{i=1}^{n} (n - d_i) \right) \left( \sum_{i=1}^{n} \frac{1}{n - d_i} \right). \]

- By the previous lemma, \( G \) has a clique of size at least \( \sum_{i=1}^{n} \frac{1}{n - d_i} \).

- By the assumption, \( \sum_{i=1}^{n} \frac{1}{n - d_i} \leq r - 1 \). Thus,

\[ n^2 \leq (r - 1) \left( \sum_{i=1}^{n} (n - d_i) \right) = (r - 1)(n^2 - 2|E|). \]

\[ \frac{n^2}{r - 1} \leq n^2 - 2|E| \quad \rightarrow \quad |E| \leq \frac{n^2}{2} \left( 1 - \frac{1}{r - 1} \right) \]
Recap

- We saw **three proofs** of Turán’s Theorem:
  - A simple proof by **induction**.
  - A proof by assigning a weight to every vertex and **manipulating** these **weights**.
  - A proof by combining the **probabilistic method** and the **Cauchy-Schwarz** inequality.

The End: Mayim Bialik

- Just like the character **Amy Farrah Fowler** that she plays in the show, **Mayim Bialik** has a Ph.D. in **neuroscience**.
  - From **UCLA**.