Combinatorial Networks
Week 4, April 2

Sperner's Theorem

- **Definition.** Let $\mathcal{F}$ be a family of subsets of $[n]$ (that is $\mathcal{F} \subset 2^{[n]}$). We call $\mathcal{F}$ an *independent system* of subsets of $[n]$, if for any two distinct $A, B \in \mathcal{F}$, we have $A \not\subset B$ and $B \not\subset A$.

  In other words, there is NO containment relationship between two sets in an independent system.

- **Definition.** (i) A *chain* of subsets of $[n]$ is a sequence of distinct sets $A_1, A_2, ..., A_k \subset [n]$ such that $A_1 \subset A_2 \subset \ldots \subset A_k$.

  (ii) A *maximal chain* is a chain with the property that no other set can be inserted in the chain.

- **Fact.** Any maximal chain $\mathcal{C}$ must like look as following:

$$\mathcal{C}: \quad \emptyset \subset \{x_1\} \subset \{x_1, x_2\} \subset \{x_1, x_2, x_3\} \subset \ldots \subset \{x_1, x_2, ..., x_n\} := [n].$$

  So it contains exactly one subset of $[n]$ of each of the possible sizes.

- **Fact.** There are $n!$ maximal chains of subsets of $[n]$.

  This is because: each maximal chain $\mathcal{C}$ (as above) defines a unique permutation $\pi : [n] \to [n]$ by $\pi(i) = x_i$. And there are $n!$ permutations.

- **Sperner's Theorem.** For any independent system $\mathcal{F}$ of subsets of $[n]$, we have $|\mathcal{F}| \le \binom{n}{\lfloor \frac{n}{2} \rfloor}$.

- First, we see this upper bound is tight, as the following independent system $\mathcal{F}$ containing exactly $\binom{n}{\lfloor \frac{n}{2} \rfloor}$ subsets: $\mathcal{F} = \{\text{all subsets of size } \lfloor \frac{n}{2} \rfloor\}$ (Why is it an independent system?)

- **Proof of Sperner's Theorem.** We will use double-counting in this proof. Given an independent system $\mathcal{F}$, we consider the set of all ordered pairs $(\mathcal{C}, A)$ such that

  (i). $\mathcal{C}$ is a maximal chain of subsets of $[n]$, and

  (ii). $A \in \mathcal{C} \cap \mathcal{F}$ is a subset of $[n]$.

  Key Observation. Any (maximal) chain can only contain at most one subset in common with any independent system! This is because that any two sets from chain will bring in the containment relationship.

  By the above observation, any maximal chain $\mathcal{C}$ contains at most one set $A \in \mathcal{F}$. Therefore,

$$\#\text{ of such pairs } (\mathcal{C}, A) = \sum_{\mathcal{C}} (\#\text{ of sets } A \in \mathcal{F} \text{ contained in maximal chain } \mathcal{C}) \le \sum_{\mathcal{C}} 1 = n!.$$

  On the other hand,

$$\#\text{ of such pairs } (\mathcal{C}, A) = \sum_{A \in \mathcal{F}} (\#\text{ of maximal chains } \mathcal{C} \text{ containing } A).$$

Let $A = \{x_1, ..., x_k\} \in \mathcal{F}$. Then any maximal chain $\mathcal{C}$ containing A must look like:

$$\emptyset \subset \{\cdot\} \subset \{\cdot, \cdot\} \subset \cdots \subset A \subset \cdots \subset [n].$$

There are $k!$ ways to form the sets between $\emptyset$ and A and there are $(n-k)!$ ways to from the sets between A and $[n]$. Therefore, the total number of maximal chains containing A is $k!(n-k)! = |A|!(n-|A|)!$. Combining the above items, we get

$$n! \geq \#\text{ of such pairs } (\mathcal{C}, A) = \sum_{A \in \mathcal{F}} (\#\text{ of maximal chains } \mathcal{C} \text{ containing } A) = \sum_{A \in \mathcal{F}} |A|!(n-|A|)!.$$

Recall that $\binom{n}{\lfloor \frac{n}{2} \rfloor}$ achieves the maximum over all binomial coefficients $\binom{n}{k}$. Therefore,

$$1 \geq \sum_{A \in \mathcal{F}} \frac{|A|!(n-|A|)!}{n!} = \sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}} \geq \sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{\lfloor \frac{n}{2} \rfloor}} = \frac{|\mathcal{F}|}{\binom{n}{\lfloor \frac{n}{2} \rfloor}},$$

which implies that $\mathcal{F}| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$. This completes the proof. ∎

- Next, we consider a nice application of Sperner's Theorem.

  Littlewood-Offord Problem. Fix a vector $\vec{\mathbf{a}} := (a_1, a_2, ..., a_n)$ with each $|a_i| \geq 1$. Let $S = S(\vec{\mathbf{a}})$ be the set of vectors $\vec{\epsilon} := (\epsilon_1, \epsilon_2, ..., \epsilon_n)$ with each $\epsilon_i = 1$ or -1 such that

  $$-1 < \vec{\mathbf{a}} \cdot \vec{\epsilon} < 1,$$

  where $\vec{\mathbf{a}} \cdot \vec{\epsilon} = \sum_{i=1}^{n} a_i \epsilon_i$. Then we always have that $|S| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$.

- *Proof.* We will reduce this problem to Sperner's theorem. For any $\vec{\epsilon} \in S$, define

  $$A_{\vec{\epsilon}} := \{i \in [n] : a_i \epsilon_i > 1\}.$$

  Then, let $\mathcal{F} := \{A_{\vec{\epsilon}} : \vec{\epsilon} \in S\}$. Therefore $|\mathcal{F}| = |S|$; also notice that $\mathcal{F}$ is a family of subsets of $[n]$. So it suffices to prove that $\mathcal{F}$ is an independent system.

  To see this, suppose for a contradiction that there are two sets $A_1, A_2 \in \mathcal{F}$ such that $A_1 \subsetneq A_2$. Let $\vec{\epsilon_i}$ be the corresponding vector in S for A_i. By definition,

  $$\vec{\mathbf{a}} \cdot \vec{\epsilon_1} = \sum_{i \in A_1} a_i \epsilon_i + \sum_{j \notin A_1} a_j \epsilon_j = \sum_{i \in A_1} |a_i| - \sum_{j \notin A_1} |a_j| = 2 \cdot \sum_{i \in A_1} |a_i| - \sum_{k=1}^{n} |a_k| \in (-1, 1);$$

  Similarly, we have

  $$\vec{\mathbf{a}} \cdot \vec{\epsilon_2} = 2 \cdot \sum_{i \in A_2} |a_i| - \sum_{k=1}^{n} |a_k| \in (-1, 1).$$

  Note that $A_1 \subseteq A_2$, so

  $$\vec{\mathbf{a}} \cdot \vec{\epsilon_2} - \vec{\mathbf{a}} \cdot \vec{\epsilon_1} = 2 \cdot \sum_{i \in A_2 - A_1} |a_i| \geq 2.$$

  But as each $\vec{\mathbf{a}} \cdot \vec{\epsilon_i} \in (-1, 1)$, we also have that $\vec{\mathbf{a}} \cdot \vec{\epsilon_2} - \vec{\mathbf{a}} \cdot \vec{\epsilon_1} < 2$. This contradiction finishes the proof that $\mathcal{F}$ indeed is an independent system.

  Then by Sperner's theorem, $|S| = |\mathcal{F}| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$. ∎

- Exercise. Find a vector $\vec{\mathbf{a}}$ in Littlewood-Offord Problem which achieves the upper bound.

Forbidden 4-cycles

- Sperner's theorem is a classic problem in extremal set theory.

  We now consider a typical problem in extremal graph theory: What is the maximal possible number of edges in an n-vertex graph G that does not contain a given "forbidden graph" F as subgraph of G?

- Choosing different forbidden graphs F (of course) will result in different answers to the above question. We look at two easy examples.

  (i). Let F be an edge. If graph G contains no copy of F, then of course graph G has no edge at all. So the answer is 0.

  (ii). Let F be a path with two edges. If graph G contains no copy of F, then any vertex of G has at most 1 incident edge. This means that the set of edges of G forms a matching! Thus, the answer (maximal number of edges) is $n/2$.

- This lecture, we focus on the case when the forbidden graph F is a cycle with four edges (or four-cycle for short). Notice that a four-cycle is exactly the complete bipartite graph $K_{2,2}$!

- We will need the following elementary but useful inequality.

  Cauchy-Schwarz inequality. For any reals $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$, we have

  $$\sum_{i=1}^{n} x_i y_i \le \sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}.$$

  Proof. Consider the inequality $\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i y_j - x_j y_i)^2 \ge 0$, which leads to

  $$2\left(\sum_i x_i^2\right)\cdot\left(\sum_i y_i^2\right) = \sum_i\sum_j (x_i^2 y_j^2 + x_j^2 y_i^2) \ge 2\sum_i\sum_j (x_i y_i)(x_j y_j) = 2\left(\sum_i x_i y_i\right)^2.$$

  After square-root, we get the desired inequality. ∎

- **Theorem.** If an n-vertex graph $G = (V, E)$ contains no copy of $K_{2,2}$ as its subgraph, then it has at most $\frac{1}{2}\left(n^{3/2} + n\right)$ edges.

- **Proof.** We use use double-counting again.

  Let S be the set containing all 3-tuples $(\{u_1, u_2\}, v)$, where three vertices u_1, u_2, v form a path $u_1 - v - u_2$ of length 2 in graph G, with v as the middle vertex.

  For any two vertices u_1, u_2 of G, there is at most one vertex v such that $(\{u_1, u_2\}, v) \in S$. To see this, suppose there are two paths $u_1 - v - u_2$ and $u_1 - v' - u_2$, then these two paths form a four-cycle $K_{2,2}$ of G, which is forbidden. Therefore, we get $|S| \le \binom{n}{2}$.

Also notice that for any two neighbors u, w of v, they form a path $u - v - w$ in G, so $(\{u, w\}, v) \in S$. This shows that $|S| = \sum_{v \in V} \binom{d(v)}{2}$, where $d(v)$ is the degree of v (i.e. the number of neighbors of v).

Thus, if using $d_1, d_2, ..., d_n$ to express the degrees of G, we get

$$\binom{n}{2} \geq |S| = \sum_{i=1}^{n} \binom{d_i}{2} \geq \sum_{i} \frac{(d_i - 1)^2}{2}.$$

Cauchy-Schwarz inequality (letting $x_i = d_i - 1$ and $y_i = 1$), together with Hands-shirking lemma, show that

$$\sqrt{\sum_{i=1}^{n} (d_i - 1)^2} \cdot \sqrt{n} \geq \sum_{i=1}^{n} (d_i - 1) = 2|E| - n.$$

Therefore,

$$2|E| \leq n + \sqrt{n} \cdot \sqrt{\sum_{i=1}^{n} (d_i - 1)^2} \leq n + \sqrt{n} \cdot \sqrt{2\binom{n}{2}} \leq n + n^{3/2},$$

which completes the proof of theorem. ∎

Forbidden triangles

- We have determined the maximal number of edges in n-vertex graphs with no copy of $K_{2,2}$ in last Lecture. Here we study the maximal number of edges in graphs with no copy of triangle. A *triangle* is a complete graph on three vertices, i.e. the K_3.
- **Definition.** Let $T(n)$ be the maximal number of edges in an n-vertex graph which doesn't contain K_3 as a subgraph.
- We first consider $T(n)$ for small values of n: $T(1) = 0$, $T(2) = 1$ which is achieved by an edge, $T(3) = 2$ which is achieved by a path of length 2, and $T(4) = 4$ which is achieved by $K_{2,2}$. Notice that all above $T(n)$ are achieved by some complete bipartite graphs.
- **Theorem 1.** For any integer $n \geq 1$, $T(n) = \lfloor \frac{n^2}{4} \rfloor$.

  (Note the floor $\lfloor x \rfloor$ of real number x denotes the largest integer which is less than or equal to x.)
- **Proof of Theorem 1.** We first show $T(n) \geq \lfloor \frac{n^2}{4} \rfloor$. Consider the complete bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$, which has n vertices and contains no copy of K_3. As $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ has exactly $\lfloor \frac{n}{2} \rfloor \cdot \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$ many edges, we see that the maximum $T(n)$ is at least $\lfloor \frac{n^2}{4} \rfloor$.

  For the proof of $T(n) \leq \lfloor \frac{n^2}{4} \rfloor$, as $T(n)$ is always an integer, it suffices to show that $T(n) \leq \frac{n^2}{4}$. In follows, we prove by induction on n that any n-vertex graph G with no copy of K_3 has at most $\frac{n^2}{4}$ edges. Base cases have been verified when $n = 1, 2, 3, 4$.

  Assume that it holds for all integers smaller than n. Consider any graph $G = (V, E)$ with $|V| = n$ and with no copy of K_3. Fix an edge $e_0 = (x, y) \in E$ and define two edge sets:

  $E_x = \{$all edges of G incident to x except $e_0\}$, $E_y = \{$all edges of G incident to y except $e_0\}$.

Let G' be a graph obtained from G by deleting vertices x, y; that is $G' = (V', E')$, where

$$V' := V - \{x, y\} \text{ and } E' := E - E_x \cup E_y \cup \{e_0\}.$$

(i) Since G' has $(n-2)$ vertices and again has no copy of K_3, by induction we get the number of edges in G' is $|E'| \le \frac{(n-2)^2}{4}$.

(ii) Note that G has no K_3, so x, y have no common neighbor in G, which means that $N_G(x) \cap N_G(y) = \emptyset$ and thus $|N_G(x)| + |N_G(y)| = |N_G(x) \cup N_G(y)| \le n$.

Combining the above properties, together with $|E_x| = |N_G(x)| - 1$ and $|E_y| = |N_G(y)| - 1$, we get that the number of edges of G is

$$|E| = |E'| + |E_x| + |E_y| + 1 = |E'| + |N_G(x)| + |N_G(y)| - 1 \le \frac{(n-2)^2}{4} + n - 1 = \frac{n^2}{4}. \blacksquare$$

- *Definition.* We say an n-vertex graph G is *extremal*, if it contains no copy of K_3 and has exactly $\lfloor \frac{n^2}{4} \rfloor$ edges.

- **Theorem 2.** For any integer $n \ge 1$, there exists a unique extremal graph on n vertices, which is the complete bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$.

  Theorem 2 tells us that there is only one graph achieving the exact maximal number. Its proof is similar to Theorem 1's. Here, we will stress the arguments we need for the structure of extremal graphs (as others are the same as in Theorem 1).

- **Proof of Theorem 2.** We prove this by induction on n. The basic case is trivial, as when $n = 2$, the extremal graph $K_{1,1}$ is unique.

  We assume that the statement holds for any smaller integer than n. Consider any extremal graph $G = (V, E)$ (with no K_3 and $|E| = \lfloor \frac{n^2}{4} \rfloor$). We will show that G must be $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$, thereby implying the uniqueness as well.

  Just similarly to the proof of Theorem 1, we fix an edge $e_0 = (x, y) \in E$ and define edge sets E_x, E_y, E' as before. And let $G' = (V', E')$, where $V' = V - \{x, y\}$. Again, we conclude that $|N_G(x)| + |N_G(y)| \le n$ (as G has no triangle). And by Theorem 1, $|E'| \le T(n-2) = \lfloor \frac{(n-2)^2}{4} \rfloor$ (as G' has no K_3). Therefore,

$$\lfloor \frac{n^2}{4} \rfloor = |E| = |E'| + |E_x| + |E_y| + 1 = |E'| + |N_G(x)| + |N_G(y)| - 1 \le \lfloor \frac{(n-2)^2}{4} \rfloor + n - 1.$$

  But $\lfloor \frac{n^2}{4} \rfloor = \lfloor \frac{(n-2)^2}{4} \rfloor + n - 1$ is an identity, which implies that all inequalities used are equalities: $|N_G(x)| + |N_G(y)| = n$ and $|E'| = \lfloor \frac{(n-2)^2}{4} \rfloor$.

  Note that G' now is an extremal graph on $n-2$ vertices. By induction, G' must be the complete bipartite graph $K_{\lfloor \frac{n-2}{2} \rfloor, \lceil \frac{n-2}{2} \rceil}$; let us say G' has a bipartition A', B', where $|A'| = \lfloor \frac{n-2}{2} \rfloor$ and $|B'| = \lceil \frac{n-2}{2} \rceil$.

  Observation: $N_G(x)$ and $N_G(y)$ are two independent set (namely having no edges at all). To see this, suppose $N_G(x)$ has an edge (u, w), then x, u, w form a K_3 in G, a contradiction.

  This observation implies that $N_G(x) - \{y\} \subset A'$ or B'; similarly, $N_G(y) - \{x\} \subset A'$ or B'.

Since $|N_G(x) - \{y\}| + |N_G(y) - \{x\}| = |N_G(x)| + |N_G(y)| - 2 = n - 2$, $|A'| + |B'| = n - 2$ and $N_G(x) \cap N_G(y) = \emptyset$, we must have that either $N_G(x) - \{y\} = A', N_G(y) - \{x\} = B'$ or $N_G(x) - \{y\} = B', N_G(y) - \{x\} = A'$. By symmetric, assume the latter case occur, by letting $A := A' \cup \{x\}$ and $B := B' \cup \{y\}$, we see now that graph G is a complete bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ with bipartite A, B. Here $|A| = |A'| + 1 = \lfloor \frac{n}{2} \rfloor$ and $|B| = |B'| + 1 = \lceil \frac{n}{2} \rceil$. ∎