CS 0368-4246: Combinatorial Methods in Algorithms (Spring 2025) March 17, 2025

Lecture 1: Finding a triangle in a graph using extermal combinatorics

Instructor: *Or Zamir* Scribes: *Liad Iluz*

1 Introduction

The first topic in the course examines whether and how the sparsity of a graph affects the complexity of solving graph problems. We begin with the following general problem:

Problem 1. *(Pattern Detection) Given a fixed-size graph H and a graph G with n vertices and m edges. Does H appear in G? In other words, is H a sub-graph of G?*

For example, if $H = K_3 = C_3$ (a triangle), we seek an efficient algorithm to determine whether G contains a triangle.

Today's lecture is about 1) Finding a triangle in a sparse graph, and 2) Extremal Combinatorics and its usages for Finding a triangle, a rectangle, a k-path and an even cycle in graphs.

2 Does G contain a triangle?

Solution 2. *Naïve solution: Check every group of three vertices. This algorithm runs in $O(n^3)$ time.*

Reminder 3. *There are two standard ways to represent edges in a graph:*

1. ***Adjacency matrix:*** *Let $G = (V, E)$ be a graph. The adjacency matrix of G is a $n \times n$-matrix, such that $A_{uv} = 1$ if $(u, v) \in E$ and 0 otherwise.*
   Space complexity: $O(n^2)$
   Checking if $(u, v) \in E$ takes $O(1)$ time.

2. ***Adjacency lists:*** *Each $v \in V$ has a list of all its neighbors.*
   Space complexity: $O(m)$
   Checking if $(u, v) \in E$ takes $O(d(u))$ time Naively, $O(\log(n))$ time using a search tree or $O(1)$ time using hash table.

Reminder 4. *Matrix Multiplication Time – Given two $n \times n$ matrices A and B, we denote by ω the best exponent such that the matrix product AB can be computed in $O(n^\omega)$ time. Currently, $2 \leq \omega \leq 2.372$. We will use fast matrix multiplication (FMM) as a black-box to solve other problems.*

Solution 5. *A graph G contains a triangle if and only if $(A^3)_{uu} = 1$ for some $u \in V$. This algorithm takes $O(n^\omega)$ time.*

Remark 6. *Note that, in general, $(A^k)_{uv}$ represents the number of paths of length k from vertex u to vertex v in the graph G. Furthermore, we could use A and A^2 to detect triangles in G.*

Question: If $H = C_4$ (a cycle of length 4), can we use the same algorithm with A^4? No! A^k represents the number of paths of length k in G, but these paths might not be simple (they may include walks). Note that C_3 is always simple. Alternatively, G contains C_4 if and only if $(A^2)_{uv} \geq 2$ for some $u \neq v \in V$.

2.1 Finding a Triangle in Sparse Graphs ($m \ll n^2$)

We aim to find an algorithm that also takes into account the number of edges, m. This way, if G has relatively few edges, the algorithm will perform better.

Solution 7. *For each edge $(u, v) \in E$, check if its endpoints form a triangle by search a common neighbor of them (i.e., $w \in (N_u \cap N_v) \setminus \{u, v\}$). This algorithm runs in $O(nm)$ time.*

Solution 8 (An $O(m^{3/2})$ Algorithm)**.** *The main idea: Define a threshold for the degree of a vertex, d, and handle vertices with degree $\leq d$ differently from those with degree $> d$. The key insight is that processing vertices with degree $\leq d$ takes little time, while sparse graphs typically contain few vertices with degree greater than d.*

In More Detail:
For each edge $e = (u, v) \in E$:

1. If $d_u > d$ and $d_v > d$, use the basic algorithm. The total number of vertices in this category is at most $\frac{\sum_v d_v}{d} = \frac{2m}{d}$, leading to an overall time complexity of $O\left(\frac{2m}{d} \cdot m\right)$.
2. If $d_u \leq d$ or $d_v \leq d$, we can check if the edge closes a triangle by examining at most d vertices. This step takes $O(md)$ time in total.

Choosing $d = \sqrt{m}$ minimizes both expressions, yielding a time complexity of $O(m^{3/2})$.

Exercises:

1. Show that we can also count the triangles in the graph in $O(n^\omega)$ time.
2. Show that we can find a triangle in $O(m^{\frac{2\omega}{1+\omega}})$ time (use both algorithms).

We now continue discussing the following questions:

1. How does the number of edges in a graph affect the existence of subgraphs from a combinatorial point of view?
2. **Sparsification:** Can we reduce a general graph to a sparse graph for certain problems while preserving the solution?
3. **Spanner:** Given an unweighted, undirected graph G and a parameter k, can we create a subgraph $G' \subseteq G$ such that:
   (a) $dist_G(u, v) \leq dist_{G'}(u, v) \leq k \cdot dist_G(u, v)$ for every $u, v \in V$; and
   (b) G' is sparse?

3 Extremal Combinatorics / Extremal Graph Theory

Definition 9 (Extremal Number of a Graph)**.** *Let H be a fixed-size graph. The extremal number of H, or the Turán number of H, denoted by $ex(H, n)$, is the maximal number of edges in a graph with n vertices that* ***does not*** *contain H as a subgraph.*

Remark 10. *The extremal number $ex(H,n)$ is well-defined, and $0 \leq ex(H,n) \leq \binom{n}{2}$.*

Theorem 11 (Mantel, 1907 [1])**.** *$ex(K_3,n) = \lfloor \frac{n^2}{4} \rfloor$*

Proof. We need to prove two things: (1) There exsits a graph with $\lfloor \frac{n^2}{4} \rfloor$ edges does not contain a triangle (i.e., K_3) as a subgraph; and (2) Every graph with $> \lfloor \frac{n^2}{4} \rfloor$ edges contains a triangle.

For the first direction, consider a complete bipartite graph with $\lfloor \frac{n}{2} \rfloor$ vertices on one side and $\lceil \frac{n}{2} \rceil$ vertices on the other side. This graph contains $\lfloor \frac{n^2}{4} \rfloor$ edges and does not have a triangle.

For the second direction, we will provide two proofs:

1. First proof — by induction on n:

   - **Base case:** $ex(K_3, 3) = 2$ and $ex(K_3, 4) = 4$ by simple checking of all possible graphs with 3 vertices and 3 edges, and with 4 vertices and 5 edges, respectively.
   - **Inductive step:** Assume that the claim holds for all values less than n. Let (u, v) be an edge in a graph with $> \lfloor \frac{n^2}{4} \rfloor$ edges. First, we have $N_u \cap N_v = \emptyset$, since otherwise a triangle would form. Therefore, $d(u) + d(v) \leq n$.

     The graph obtained by removing u, v, and all edges incident to them is a graph with $n-2$ vertices and no triangle, since otherwise we would have a triangle in G. Notice that at most $n-1$ edges are deleted from the original graph. From the induction hypothesis:

     $$ex(K_3, n) \leq ex(K_3, n-2) + (n-1) \leq \lfloor \frac{(n-2)^2}{4} \rfloor + n - 1 = \lfloor \frac{n^2}{4} \rfloor.$$

2. Second proof — using Jensen's inequality: We saw that for every edge $(u, v) \in E$, $d(u) + d(v) \leq n$. Therefore,

   $$m \cdot n \geq \sum_{(u,v)\in E} (d(u) + d(v)) = \sum_{v\in V} d(v)^2 = n \cdot \frac{1}{n} \sum_{v\in V} d(v)^2 \geq_{\text{Jensen}} n \cdot \left(\frac{1}{n} \sum_{v} d(v) \right)^2 = \frac{4m^2}{n}.$$

   So, we have $m \leq \frac{n^2}{4}$.

$\square$

Theorem 12 (Turan, 1941 [2])**.** *$ex(K_{r+1}, n) \leq \left(1 - \frac{1}{r}\right) \cdot \frac{n^2}{2}$*

Proof. First direction ($\leq$): As done previously, we divide the n vertices into r equal groups and connect any pair of vertices that are not in the same group. By the pigeonhole principle, there is no K_{r+1} in this graph (See figure 1). Moreover, the number of edges is

$$\frac{n^2}{2} - r \cdot \frac{(n/r)^2}{2} = \left(1 - \frac{1}{r}\right) \cdot \frac{n^2}{2}.$$

The second direction is given as an exercise (it can also be done by induction). $\square$

Theorem 13 (Erdős-Stone-Simonovits [3])**.** *$ex(H, n) = \left(1 - \frac{1}{\chi(H)-1} + o(1)\right) \cdot \frac{n^2}{2}$, where $\chi(H)$ is the chromatic number of H.*

This is given without a proof (the first direction is similar).

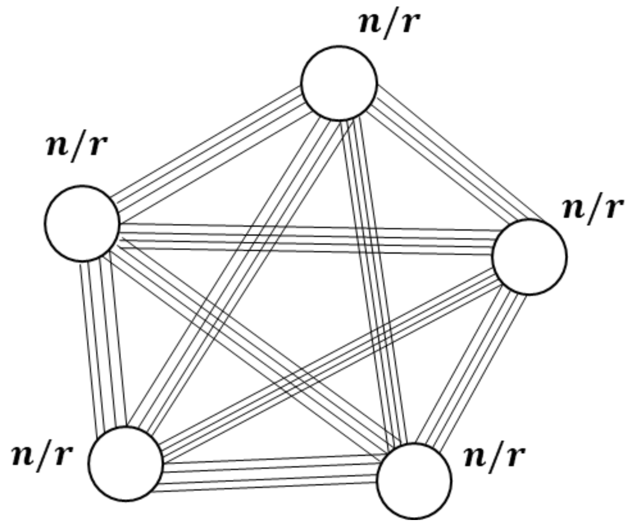

Figure 1: The first direction of the proof of Turan's theorem

3.1 Does G contain a path of length k? Rectangle? Even cycle?

We may notice that if H is a bipartite graph, the last theorem does not provide useful information, as it would give a condition of $o(1) \cdot \frac{n^2}{2}$ edges. We now want to explore further for P_k, a simple path of length k, and for C_{2k}, a cycle of even length.

Theorem 14. $ex(P_k, n) \leq kn$

Lemma 15. *If a graph has $m \geq d \cdot n$ edges, then it contains a subgraph with a minimum degree of at least d.*

Proof. We prove this by greedy construction. Unless the graph contains a vertex with degree less than d, we repeatedly remove such vertices along with their incident edges. We need to show that this process will end with a non-empty graph. Initially, the average degree is $\frac{2m}{n} \geq 2d$. We claim that the average degree will always remain $\geq 2d$. Therefore, the graph will never become empty. Indeed, in every deletion, the number of vertices decreases by 1, and the sum of all degrees decreases by at most $2d$. □

Proof. (of the theorem) From the previous lemma, any graph with at least kn edges contains a subgraph with a minimum degree of at least k. We now start with some vertex in this subgraph and proceed along a path of length $k-1$, where at each step, we choose a neighbor that has not yet appeared on the path. □

Theorem 16 (Zarnkevitch)**.** $ex(C_4, n) = \Theta(n^{3/2})$

Proof. First direction: We will show that every graph with $O(n^{3/2})$ edges contains C_4. A crucial observation is that if the number of 2-paths, $\#P_2$, is at least n^2, then the graph contains a C_4 (this follows from the pigeonhole principle, which guarantees that there are at least two distinct 2-paths between the same pair of vertices). Therefore, it suffices to show that if $m > 2n^{3/2}$, then $\#P_2 \gg n^2$. Let's compute:

$$\#P_2 = \sum_v \binom{d(v)}{2} \geq n \cdot \binom{\frac{1}{n}\sum_v d(v)}{2} = n \cdot \binom{\frac{2m}{n}}{2} \approx \frac{2m^2}{n},$$

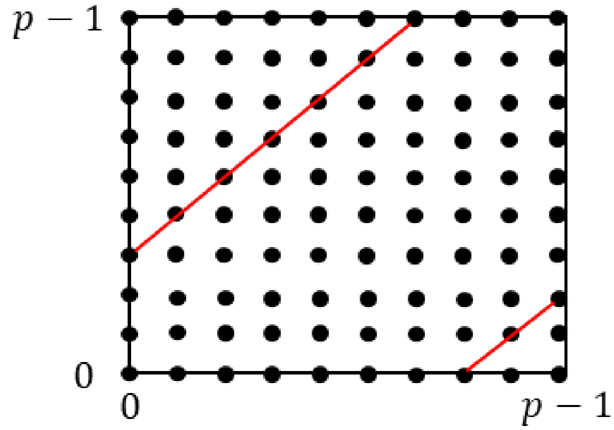

Figure 2: A line in the plane $\mathbb{Z}_p^2$

where the first step follows from considering the middle vertex in each 2-path, and the second step uses Jensen's inequality.

Therefore, if

$$\frac{2m^2}{n} > n^2 \iff m^2 > \frac{n^3}{2} \iff m > \frac{n^{3/2}}{\sqrt{2}},$$

then G contains a C_4 as a subgraph.

Second direction: We will show that $ex(C_4, n) \geq \Omega(n^{3/2})$, i.e., we will construct a graph with $\Omega(n^{3/2})$ edges that does not contain a C_4. Let p be a prime number. Consider the plane $\mathbb{Z}_p^2$ and define all lines in this plane as follows (See figure 2):

$$\forall y_0, m \in \mathbb{Z}_p : L_{y_0,m} := \{(x, y_0 + mx) \mod p : x \in \mathbb{Z}_p\}.$$

From this definition, we get two properties of the lines:

1. For every y_0, m: $|L_{y_0,m}| = p$.
2. If $(y_1, m_1) \neq (y_2, m_2)$, then $|L_{y_1,m_1} \cap L_{y_2,m_2}| \leq 1$.

The first property is clear. For the second property, consider $(x, y) \in L_{y_1,m_1} \cap L_{y_2,m_2}$. From the definition of the lines:

$$y_1 + xm_1 = y_2 + xm_2 \text{ (mod p)} \implies y_1 - y_2 = x(m_2 - m_1) \text{ (mod p)}$$

Now, if $m_1 = m_2$, then $y_1 = y_2$, meaning either both lines are the same line or parallel lines. Assuming $(y_1, m_1) \neq (y_2, m_2)$, we are in the later case, so $|L_{y_1,m_1} \cap L_{y_2,m_2}| = 0$. Otherwise, $m_1 \neq m_2$, so there exists an inverse to $m_2 - m_1$ in the field $\mathbb{Z}_p^*$, and thus $x = (m_2 - m_1)^{-1}(y_1 - y_2)$. Hence, there is exactly one common point on the two lines.

Now, we will construct the following graph: G is a bipartite graph with p^2 vertices on the left side (each representing a line in the plane, $L_{y,m}$) and p^2 vertices on the right side (each representing a point in the plane (x, y)). We create an edge between a line and all the points that lie on that line, i.e., $(L_{y,m}, (x, y)) \in E \iff (x, y) \in L_{y,m}$ (See figure 3). This results in a graph G with $n = 2p^2$ vertices and $m = p^3 = \Theta(n^{3/2})$ edges.

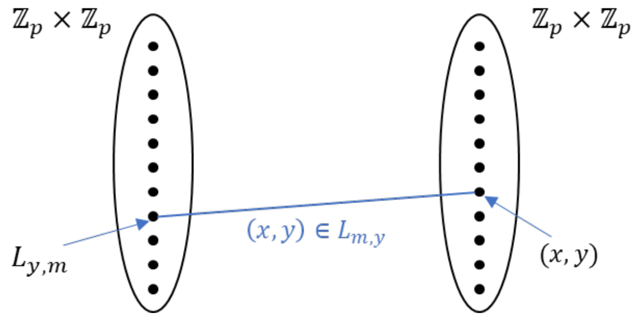

Figure 3: The lower bound – graph construction

Observe that a bipartite graph contains a C_4 if and only if there are two vertices with at least two common neighbors. From the second property and the observation above, we conclude that G does not contain a C_4. □

We can now describe an algorithm to find C_4 in a graph with $O(n^2)$ time:

Algorithm 17. *Given a graph G.*
1. Initialize a 2-dimensional array of size $n^2/2$ with zeros.
2. For every $v \in V$:
2.1. For every pair of neighbors of v : u, w:
2.1.1 If $A_{u,w}$ is marked, return YES.
2.1.2 Otherwise, mark $A_{u,w}$.
3. Return NO.

By the first part of the last proof, after at most n^2 iterations, we will finish. Therefore, this is $O(n^2)$-time algorithm.

Theorem 18 (Bondy-Simonovitz [4])**.** $ex(C_{2k}, n) \leq k \cdot n^{1+\frac{1}{k}}$

We will not prove this theorem. As of today, we do not know whether this upper bound is tight, except for some constructions for C_4, C_6, and C_{10}.

Theorem 19 (Moore)**.** *If a graph has $m \geq 100 \cdot n^{1+\frac{1}{k}}$ edges, then it contains a cycle of length at most $2k$ (i.e., the girth of the graph is at most $2k$).*

Proof. By the previous lemma, there exists a subgraph with a minimal degree of $d = 100 \cdot n^{\frac{1}{k}}$ (where $d \geq 100$). Consider some vertex in this subgraph, and look at all paths starting from this vertex up to depth k (for example, using BFS). Observe that if we revisit the same vertex, we have found a cycle of length at most $2k$. Otherwise, all the vertices we encounter are distinct. Therefore, the number of vertices in this subgraph will be at least:

$$1 + (d-1) + (d-1)^2 + \cdots + (d-1)^k = \frac{(d-1)^{k+1} - 1}{d-1-1} \geq (d-1)^k - 1 \geq \left(\frac{d}{2}\right)^k$$

But then,

$$\left(\frac{d}{2}\right)^k \geq 50 \cdot n > n,$$

which leads to a contradiction. Hence, the graph must contain a cycle of length at most $2k$. □

Back to the spanner question that we introduced earlier, we can now show an algorithm to construct a sparse subgraph that preserves distances up to a factor of k:

Algorithm 20. *Given a graph G and a parameter k. Go over every edge $(u, v) \in E$, and add it to the new (sparse) graph G' iff $dist_{G'}(u, v) > k$ at that moment.*

At the end of the algorithm, for every $(u, v) \in E$, there will be an alternative path between u and v with length at most k, and essentially all such paths will have length at most k. Therefore, the condition of preserving the path lengths, i.e., $dist_{G'}(u, v) \le k \cdot dist_G(u, v)$ for every $u, v \in V$, holds. Moreover, G' does not contain a cycle of length at most $k+1$, since we would not have added the edge that closes such a cycle to G'. Hence, by Moore's theorem, we have:

$$e(G') \le 100 \cdot \frac{k+1}{2} \cdot n^{1+\frac{1}{(k+1)/2}} \le O(n^{1+\frac{2}{k+1}}).$$

References

[1] W. Mantel, Problem 28, In Wiskundige Opgaven 10 (1907), 60–61.

[2] P. Tur´an, Eine Extremalaufgabe aus der Graphentheorie, Mat. Fiz. Lapok 48 (1941), 436–452.

[3] P. Erdős and A.H. Stone, On the structure of linear graphs. Bulletin of the American Mathematical Society. 52 (12), (1946), 1087–1091.

[4] J.A. Bondy, and M. Simonovits, Cycles of even length in graphs, Journal of Combinatorial Theory, Series B, 16 (2),(1974) 97–105.