Lecture 6: Graph Properties

Rajat Mittal[*]

IIT Kanpur

In this section, we will look at some of the combinatorial properties of graphs. Initially we will discuss independent sets. The bulk of the content will be about properties like coloring and matching.

A vertex coloring of a graph (we will only worry about vertex coloring in this course) is an assignment of colors to the vertices such that no two adjacent vertices have the same color.

Coloring is useful in many contexts: Assigning students to exam rooms such that no two friends sit together, assigning frequencies to radio towers such that no two towers in the range of each other should interfere, Sudoku and coloring the map so that no two neighbouring states get the same color.

Exercise 1. Show that all these examples are coloring problems, find the vertices and edges of the graph whose coloring represents the problem.

One of the natural strategy could be to first look at the biggest subset of vertices which can be colored by the same color. Subsets, which can be colored by the same color, are known as independent sets. Let us study them first.

1 Independent sets

Suppose we are given a graph where vertices correspond to students in the class and edges correspond to friendships. What is the maximum number of students we can put in the exam room, if no two friends should be in the room?

If no two friends are allowed to sit in the same exam room, there will be no edges between the people sitting in the exam room. This constraint defines the concept of an "independent" set.

Formally, an *independent set* of a graph $G = (V, E)$ is a subset S of vertices V, s.t., no two elements u, v of S have an edge between them. In mathematical notation,

$$u, v \in S \Rightarrow (u, v) \notin E.$$

It is also called a *stable set*. The stability number, denoted by $\alpha(G)$, of a graph G is the maximum possible size of an independent set in the graph.

Exercise 2. What is the stability number of the graph given in Fig. 1?

We can also define an associated term, *maximal independent set*, as an independent set where adding any other vertex will make the subset *not* independent.

Exercise 3. Construct a connected graph where there is a maximal independent set whose cardinality is lesser than the stability number.

Any "star" is an easy example. Eg. $\{(4, 1), (4, 2), (4, 3)\}$

Remember, we defined the *complement* of a graph G by taking edge (u, v) if and only if (u, v) is not an edge in G. The complement of a graph G is denoted by $\bar{G}$. So,

$$E_{\bar{G}} = \{(u, v) : \ (u, v) \notin E_G\}.$$

Exercise 4. What are the number of edges in $\bar{G}$ if there are m edges and n vertices in G.

[*] Thanks to Nitin Saxena for his notes from the previous iteration of the course.

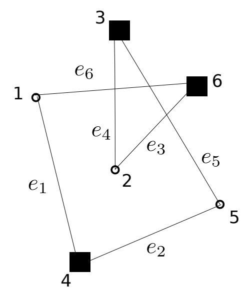

Fig. 1. Independent set in a graph. Is that the biggest possible?

Suppose S is an independent set in G. Consider the subgraph of $\bar{G}$ on vertices of S; it is going to be a complete graph. Such subgraphs are called *cliques*.

More formally, a *clique* S of a graph G is the subset of vertices where there is an edge between every possible pair of vertices. In other words, G is a complete graph when restricted to the vertices of S.

From the definitions, it is clear that an independent set in the graph is same as a clique in the complement graph.

Another connected notion is of a *vertex cover*. A vertex cover S is a subset of the vertex set, s.t., every edge has at least one vertex in common with S.

Exercise 5. Show that S is a vertex cover iff $V - S$ is an independent set.

So finding the best vertex cover (minimum) is same as finding the best independent set (maximum).

It is easy to find an approximate vertex cover. We can take a *greedy* approach. Pick an edge (include both the vertices of the edge in the vertex cover) and delete all the edges connected to it. In the remaining graph, again pick an edge and so on.

Exercise 6. This algorithm gives a vertex cover which is at most *twice* the size of the optimal vertex cover.

Let $e_1, \ldots, e_k$ be the edges picked. Any vertex cover has to pick, for every e_i, at least one of the end-points. Thus, every vertex cover has size at least $k/2$.

We believe that finding the maximum independent set in a graph is a hard problem. If there is an efficient algorithm for it then there will be efficient algorithms for a large class of very interesting problems (eg. NP-complete ones).

Not just that, finding even an approximation algorithm for independent set is really hard. Convince yourself that if you use the approximation algorithm for vertex cover and take its complement; the result need not be a good independent set.

2 Coloring

Consider the following problem. For an examination, we need to put students in classrooms such that no pair of friends sit in the same classroom. We are given the social graph of the class, with students as vertices and edges representing friendships. What is the minimum number of classrooms needed for the examination?

Look at any valid assignment. Assign a distinct color for every classroom and then color a vertex corresponding to its classroom. By this process, no two adjacent vertices will get the same color. This is called a "coloring" of the graph.

Note 1. The name of the colors are immaterial.

To define it precisely, given a graph G, a *valid coloring* (or just coloring) is a map from a set of colors to V, s.t., no two vertices of the same color are adjacent. The minimum number of colors needed to have a valid coloring of a graph G is known as its *chromatic number* $\chi(G)$.

The above question about examination classrooms can be reformulated as– what is the chromatic number of the social graph. There are many other applications of coloring,

– Color the map of countries so that no two countries with shared border have the same color.
– Schedule the examinations so that no student has two exams in the same slot.

Exercise 7. Formulate all the above questions as graph coloring problems.

(1) Countries become vertices. Adjacent ones are joined by an edge. (2) Courses become vertices. A & B are adjacent if there is a student in both these courses.

Exercise 8. What is the chromatic number of a cycle on n vertices?

Odd cycle needs 3 colors. Even ones need only 2.

Exercise 9. Show that bipartite graphs are the graphs with chromatic number less than equal to 2.

For a graph coloring, the first thing to notice is that every color class (set of vertices with that color) form an independent set. So a coloring is equivalent to partitioning the vertex set into *disjoint* independent sets.

A natural approach to color a graph seems to be, find a maximum independent set, color it with first color. Look at the remaining graph, find the maximum independent set in the new graph, color with second color and so on.

Exercise 10. Show a counterexample, i.e., a graph where this strategy fails.

Though, we can still find some useful connections between $\alpha(G)$ and $\chi(G)$.

Theorem 1. *Given a graph G with n vertices. Then $\alpha(G)\chi(G) \geq n$.*

Proof. Let us say that the optimal coloring divides the vertex set V into color classes $V_1, V_2, \cdots, V_k$ ($\chi(G) =: k$). It means that two vertices in the same V_i have same color and two vertices in different V_i's have different colors. Then,

$$n = \sum_{i=1}^{k} |V_i|.$$

But every V_i is an independent set and hence $|V_i| \leq \alpha(G)$ for all i. So,

$$n \leq k \cdot \alpha(G) = \chi(G)\alpha(G)\,.$$

□

Exercise 11. Construct a connected graph with at least 6 vertices, s.t., $\chi(G)\alpha(G) = n$.

6-cycle.

An upper bound on chromatic number can be given by the degree.

Theorem 2. *If G has maximum degree k, then $\chi(G) \leq k + 1$.*

Moreover, if G is connected and there is at least one vertex with degree strictly less than k then $\chi(G) \leq k$.

Proof. We will first prove that $k+1$ colors suffice to color a graph with degree k. Let us apply induction on number of vertices in the graph. For the base case, clearly if the number of vertices are less than or equal to $k+1$ then graph can be colored with $k+1$ colors.

For the general case, consider any particular vertex v. Let G' be the graph obtained by deleting v and all edges incident on v. G' has maximum degree k and has less number of vertices. So G' can be colored with $k+1$ colors.

Now consider the neighbors of v. There are k of them, pick the $(k+1)$-th color for v. Hence, G can be colored with $k+1$ colors.

For the second part, we will again use induction. There exists a vertex with degree less than k, say v. Remove v and all edges incident on v. Suppose we get connected components $H_1, H_2, \cdots, H_\ell$. All H_i's have to be connected to v, since G is connected.

Exercise 12. At least one vertex in every H_i has degree less than k.

The vertex connected to v.

Applying induction hypothesis on all H_i's we get a coloring for all of them. Since they are disconnected, the entire coloring is consistent. Now, for v, there are at most $k-1$ neighbors. So, the remaining color can be used to color v. Hence proved.

□

Exercise 13. Construct a graph with maximum degree d which is not colorable using d colors.

A triangle.

Exercise 14. Give a graph $G=(V,E)$ whose degree is $O(|V|)$ but chromatic number is constant.

3 Matching

Suppose you are the organizer of a friendly tennis tournament between India and Pakistan. You have been provided with a graph with vertices as the participants. There are edges between the participants if there is a match possible between them (the players have the same level). How many maximum matches can you hold simultaneously?

This is called a matching problem. You want to match participants in pairs so that,

- all pairs are disjoint and
- number of pairs are maximized.

For the following discussion on matching, we will assume that the graph is bipartite. The two set of vertices are denoted by X and Y, s.t., $|X| \leq |Y|$.

A *matching* M in a bipartite graph $G=(X \cup Y, E)$ is a subset of edges, s.t., no two edges share a vertex.

There are two closely related concepts, a *maximum* matching and a *maximal* matching.

- A maximum matching is a matching M, s.t., for any matching N in G, $|M| \geq |N|$.
- A maximal matching is a matching M, s.t., no extra edge from graph G can be added to it without violating the matching property.

A maximum matching is maximal by definition. Is the opposite true?

Exercise 15. Is the matching shown by bold edges maximum (Fig. 2)? Is it maximal?

No and Yes

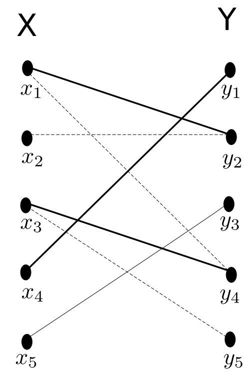

Fig. 2. An alternating path. Thick edges are the edges of existing matching.

It is not a maximum matching because there exists a matching with 5 edges. How can we find it out?

Suppose there is a path which starts from a non-matched vertex, alternates between a matched and a non-matched edge and finishes at a non matched vertex. The path $x_2, y_2, x_1, y_4, x_3, y_5$ (in the Fig. 2) is such an example. By putting $(x_2, y_2), (x_1, y_4), (x_3, y_5)$ in the matching and removing $(x_1, y_2), (x_3, y_4)$ we can increase the size of the matching.

This idea gives us the following definition. An *alternating path* in a graph, with a given matching M, is a path where:

- the first (in X) and the last (in Y) vertex are not matched, and
- the edges alternate between unmatched and matched edges.

You should convince yourself that if such a path exist then we can increase the size of matching. You just need to show that picking the alternate edges preserves the matching property. Surprisingly, the converse is true.

Theorem 3. *A matching is maximum iff there are no alternating paths.*

Proof. As noticed, if there is an alternating path P then the matching M is not maximum. (Note: We change the matching to $M' := (M \setminus P) \cup (P \setminus M)$; also denoted by the symmetric difference $P \Delta M$.)

We will now prove that if a matching M is not maximum then there is an alternating path.

Suppose M' is a maximum matching. Consider the graph G' which has only edges from M and M'. If M and M' share an edge, we will keep *both* edges from them.

Clearly every vertex has degree at most 2. Hence, the connected components in G' will look like isolated vertex, cycle or path. We will not worry about isolated vertices, as they will not appear in any matching.

Notice that any path/cycle in the bipartite graph G' has edges alternating between edges from M and edges from M' (A vertex has degree 1 in a matching). This shows that any cycle will be of even length and those can be ignored (same number of edges from M and M').

Let us focus on paths. We know that the path will have alternating edges from M and M'. Since the number of edges in M' is greater than M, there exists a path where number of edges of M' are higher than the number of edges from M. We know that the path is alternating, so the difference in number of edges from M and M' should be exactly one.

Exercise 16. Prove that any connected component C, where number of edges from M are less than the number of edges from M', is an alternating path for M.

Difference in two types of edges in C is 1 (it is a path with alternating edges). The first and the last edges must belong to M'.

Thus, at least one alternating path for M will exist in G' (& hence in G). So, if a matching is not maximum, there exists an alternating path. $\square$

4 Complete matching

Let M be a matching in $G = (X \cup Y, E)$ with $|X| \leq |Y|$. If all vertices of X are included in M then it is called a *complete matching*[1]. For the above example of tennis matches, if there is a complete matching, then all participants (of country X) can play at the same time with someone of their own level.

Though it is not necessary that a graph has a complete matching. You might notice that if there is an isolated vertex (degree zero vertex) then there is no such matching. The guess would be that if all vertices have high enough degree then there should be a complete matching.

Exercise 17. Find a bipartite graph such that all degrees are higher than 3, which does not have a complete matching. Show that 3 can be replaced by any constant.

Consider the union of two complete bipartite graphs $G := K_{d,d+1} \sqcup K_{d+1,d}$. This is a bipartite graph with $|X| = |Y| = 2d + 1$, with degree at least d, but with no complete matching.

Given a subset S of vertices X, define $N(S)$ to be the set of vertices in Y adjacent to S.

$$N(S) := \{v \in Y : \exists\, u \in S : (u, v) \in E\}\,.$$

The set $N(S)$ is called the *neighborhood* of S. Suppose the cardinality of $N(S)$ is smaller than $|S|$ for a graph G. Then we would not be able to match every vertex in S.

In other words, if there exists a subset S, s.t., $|N(S)| < |S|$ then there is no complete matching. We would like to show that the converse is also true.

The converse of this condition, $|N(S)| \geq |S|$ for all subsets S of X, is called *Hall's condition*. We want to show, if Hall's condition is true then there is a complete matching.

The idea is: if Hall's conditions are satisfied and the matching is not maximum, then we can always grow the matching (there exists an alternating path).

Theorem 4 (Hall's marriage theorem, 1935). *Let $G = (X \sqcup Y, E)$ be a bipartite graph. There is a complete matching in G iff for all $S \subseteq X$, $|N(S)| \geq |S|$.*

Proof. Clearly, if there is a complete matching then the size of the neighborhood $N(S)$ for any $S \subseteq X$ has to be larger than or equal to the size of S.

For the opposite direction, suppose that Hall's conditions are satisfied and M is a maximum matching which is not complete. We will show the existence of an alternating path in G with respect to M.

Consider any vertex x_0 in X which is not matched. $N(\{x_0\})$ will have at least one element, say y_1. If y_1 is not matched we are done else it is matched to x_2. For the set $\{x_0, x_2\}$, again we can find a y_3 (not equal to y_1) which is connected to either x_0 or x_2.

Note 2. y_3 need not necessarily connect to x_2. In other words, x_0, y_1, x_2, y_3 need not be an alternating path

Still, we keep growing the set using Hall's condition and matching with $x_4, y_5, \cdots$. Continuing this way, we should reach a vertex y_r which is not matched.

The alternating path can be constructed by tracing back. y_r should be connected to some x_i. We know that x_i matches to y_{i-1}. By the process of selecting x_j's and y_j's, y_{i-1} is connected to some x_t where $t \leq i-2$. This gives the alternating path:

$$y_r, x_i, y_{i-1}, x_t, y_{t-1}, \cdots, x_0\,.$$

But there cannot be an alternating path for a maximum matching by Thm. 3. This contradiction implies that M is a complete matching. $\square$

[1] If $|X| = |Y|$ then it is also called a *perfect matching*.

5 Extra reading: planarity [2]

A *planar representation* of a graph G is a drawing of the graph, say, on a piece of paper such that no two edges intersect each other except at the end points. Take a look at two representations of the graph K_4.

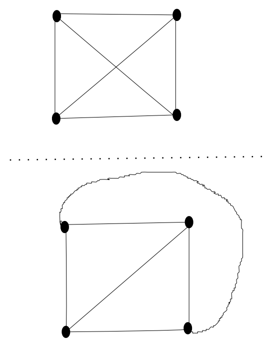

Fig. 3. Two different representations of K_4. Second one is planar.

A graph G is called *planar* if it has a planar representation. Notice that a planar graph can also have non-planar representations. To show that a graph is planar, we can just show a planar representation. But showing that a graph is non-planar takes a lot of effort. Let us take an example.

Theorem 5. *The graph K_5 is non-planar.*

Proof. Any planar drawing of a graph divides the plane into regions. Look at the following example of K_4.

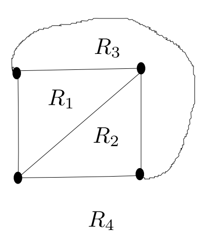

Fig. 4. Regions in a graph.

[2] This section is taken from the book by Rosen [1].

In any planar drawing of the graph K_5, pick 4 vertices and call them v_1, v_2, v_3, v_4. Since $(v_1, v_2), (v_2, v_3), (v_3, v_4)$ and (v_4, v_1) are all connected. They will divide the plane into two regions. We can call them *inside* and *outside*.

Case 1: v_5 falls in the region *inside*. Then both the edges (v_1, v_3) and (v_2, v_4) have to fall in the region *outside*. But then they will have to cross each other.

Case 2: v_5 falls in the region *outside*. The proof in this case is similar and left as an exercise. □

Exercise 18. Show that $K_{3,3}$ is non-planar.

Consider a planar drawing of the graph $K_{3,3} = (X \sqcup Y, E)$ with $X = [3], Y = \{4, 5, 6\}$. Now $(1, 4, 2, 5, 1)$ forms a cycle C which divides the plane into two regions– call them *inside* and *outside*. Since $(3, 6)$ is an edge, we have to put them in the same region.

Case 1: Vertices $3, 6$ fall in the region *inside*. First, place 3 and divide the inside into two more sub-regions. Next, placing 6 becomes impossible as $(6, 1), (6, 2)$ are edges.

Case 2: Vertices $3, 6$ fall in the region *outside*. First, place 3 and divide the outside into two more sub-regions. Next, placing 6 becomes impossible as $(6, 1), (6, 2)$ are edges.

If a graph G has $K_{3,3}$ or K_5 as a subgraph, then, clearly, G is non-planar.

An *elementary subdivision* is an operation on a graph G whereby we can replace an edge (u, v) by two edges $(u, w), (w, v)$ by adding a new vertex w.

Graphs G_1, G_2 are called *homeomorphic* if they can be obtained from a third graph G by applying a sequence of elementary subdivisions.

Exercise 19. If G is planar then G_1, G_2 are also planar.

A planar representation of G immediately gives one for G_1, G_2.

Eg. 10-vertex Petersen graph's non-planarity.

Surprisingly, it can be shown, with much work, that: a graph is non-planar iff it has a subgraph which is homeomorphic to $K_{3,3}$ or K_5. Interested readers can look at *Kuratowski Theorem* (1930). This gives an efficient algorithm for planarity testing of graphs!

In the above proofs of non-planarity, we studied the different regions 'created' by a representation of the graph. Euler showed that for a graph G, any planar representation have the same number of regions, and this number is related to vertices and edges in a simple way.

Theorem 6 (Euler's formula, 1752). *Let G be a connected planar graph with n vertices and m edges. The number of regions r in any planar representation is $m - n + 2$.*

Proof. Let us keep a particular planar representation of G in mind. We are going to construct this representation by adding one edge at a time. We will start with any single edge– call this base graph G_1.

Given G_i, look at a new edge e ($e \notin E(G_i)$) which has at least one vertex in G_i. e exists at every step because the graph G is connected. If both the endpoints of e are already in G_i, to obtain G_{i+1}, we just need to draw the edge. Otherwise, to obtain G_{i+1}, draw the edge and the additional vertex too.

Suppose r_i, e_i, v_i be the number of regions, edges and vertices respectively in the graph G_i. We will show by induction that $r_i = e_i - v_i + 2$.

Exercise 20. Show the base case.

In G_1, $r_1 = 1, e_1 = 1, v_1 = 2$. So, $r_1 = e_1 - v_1 + 2$.

As mentioned above, there can be two cases when adding a new edge e to G_i.

Case 1: Both vertices of e are already present in G_i. Then they should be in the same region (otherwise there will be a crossing). By connecting those two vertices, we have divided the region into two regions. So $r_{i+1} = r_i + 1, e_{i+1} = e_i + 1, v_{i+1} = v_i$ and Euler's formula continues to hold.

Case 2: Only one vertex of e is present in G_i. In this case the new edge does not make a new region. So $r_{i+1} = r_i, e_{i+1} = e_i + 1, v_{i+1} = v_i + 1$ and again Euler's formula holds.

□

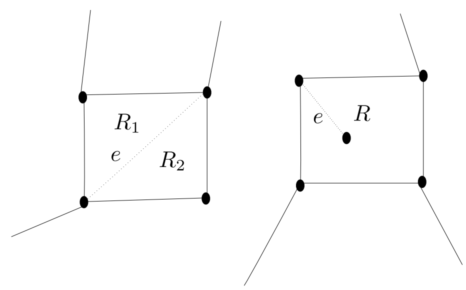

Fig. 5. Two different cases of adding an edge.

For a tree Euler's formula implies $m = n - 1$.

A very interesting fact is known about any planar graph– It can always be colored by 4 colors (Appel & Haken, 1976). This is known as 4-*color theorem* and the proof of it required a lot of effort (it is a computer-assisted proof!). If you are interested please read more about the 4-color theorem on the internet.

References

1. K. H. Rosen. Discrete Mathematics and Its Applications. *McGraw-Hill*, 1999.
2. N. L. Biggs. Discrete Mathematics. *Oxford University Press*, 2003.