Math 3322: Graph Theory[1] Mikhail Lavrov

Lecture 23: Cliques and independent sets

November 5, 2024 *Kennesaw State University*

1 Defining cliques and independent sets

1.1 Definitions

Cliques and independent sets are closely related objects in graphs.

A **clique** in a graph G is a subgraph isomorphic to K_k for some k. Equivalently, if we think about it as a set of vertices, it is a subset $S \subseteq V(G)$ such that any two vertices $u, v \in S$ are adjacent.

For any vertex v, $\{v\}$ is a one-vertex clique. For any edge uv, $\{u, v\}$ is a two-vertex clique. However, with more vertices, the number of constraints grows rapidly: large cliques are rare in most graphs, and hard to find. For this reason, there is an associated maximization problem: if I give you a graph G, what is the largest clique you can find in it? The number of vertices in a largest clique in G is denoted $\omega(G)$: the **clique number** of G.

An **independent set** in a graph G is a subset $S \subseteq V(G)$ such that there is *no* pair of adjacent vertices $u, v \in S$. Once again, there is an associated maximization problem: if I give you a graph G, what is th largest independent set you can find in it? The number of vertices in a largest independent set in G is denoted $\alpha(G)$: the **independence number** of G.

These definitions are very similar: the only change is whether we want vertices to be adjacent or non-adjacent. As a result, cliques in a graph G correspond exactly to independent sets in the complement graph $\overline{G}$ (where adjacent vertices become non-adjacent, and non-adjacent vertices becomea adjacent). Similarly, independent sets in G correspond exactly to cliques in $\overline{G}$. In particular, $\alpha(G) = \omega(\overline{G})$ and $\omega(G) = \alpha(\overline{G})$.

There are also two connections to parameters we have previously studied.

- A subset $S \subseteq V(G)$ is an independent set if and only if its complement $V(G) - S$ is a vertex cover. For S to be an independent set, no edge can have both endpoints in S. For $V(G) - S$ to be a vertex cover, each edge must have at least on endpoint in $V(G) - S$. These say the same thing.

  (This statement sounds superficially similar to the relationship between independent sets and cliques, since both have the word "complement". Don't be confused by this; in one, we are taking the complement of the graph, and in the other, we are taking the complement of the set. Think through the definitions when you use these properties!)

  Recall that $\beta(G)$ denotes the number of vertices in the smallest vertex cover. When S is as large as possible, $V(G) - S$ is as small as possible; therefore $\alpha(G) + \beta(G)$ is always equal to $|V(G)|$.

[1] This document comes from an archive of the Math 3322 course webpage: http://misha.fish/archive/3322-fall-2024

- Conceptually, independent sets and matchings are related. Independent sets are sets of vertices that share no edges; matchings are sets of edges that share no vertices.

  There is no direct relationship between independent sets and matchings in a graph (except through vertex covers). However, the similarity in definitions is the reason why the notation is similar. We write $\alpha(G)$ for the size of a largest independent set, and $\alpha'(G)$ for the size of a largest matching.

1.2 Examples

In principle, we don't *need* three parameters $\alpha(G), \beta(G), \omega(G)$. We could express all three of them in terms of just one: $\alpha(G), |V(G)| - \alpha(G), \alpha(\overline{G})$. However, in applications, sometimes one of them is easier to think about than the other.

In the first lecture, we saw an application of independent sets to tiling problems. Here are a few others.

Scheduling and interval graphs. Suppose we are trying to schedule events. This is an easy task if none of the events overlap, but of course they might. Let's say that the i^{th} event starts at time a_i and ends at time b_i; two events conflicts conflict if the intervals $[a_i, b_i]$ and $[a_j, b_j]$ intersect.

We can draw a graph G representing the conflicts: it has a vertex for every event, and an edge for every two events with a conflict between them. Such a graph is often called an **interval graph**. Both cliques and independent sets are of interest to us with interval graphs:

- A clique in G is a set of events among which any two are in conflict. This happens if there is at least one moment time at which all the events in the clique are happening simultaneously.
- An independent set in G is a set of events among which none conflict. If you are looking for a set of events that can all be scheduled in the same location, or a set of events that one person can attend, you are looking for an independent set.

Friend groups. Cliques are common to encounter when we are looking at a graph representing people that know each other—for example, the graph of Facebook, where vertices are Facebook accounts, and edges are friendships. The graph-theoretic use of the word "clique" comes from the usual sense of "clique" as it occurs in this application: a clique in the Facebook graph is a set of people that are all friends on Facebook.

It is not surprising to see large cliques in this graph; just think of how many times you see "XX mutual friends" on Facebook! Often, these cliques correspond to a club or event in real life, so advertisers might be interested in finding them… but let's not talk about such dark arts.

For graph theorists, Facebook cliques are interesting for a different reason: modeling. How would you create a graph that simulates the Facebook graph? The easiest way is to choose edges at random: maybe you have a million vertices, and each vertex is joined to a random set of (say) 100 other vertices. However, cliques tell us that this is a bad model: the real Facebook graph has a much higher clique number than such a random graph. More sophisticated models are necessary.

(The resulting graphs could be used, for example, to study the speed of disease propagation, assuming that an illness is more likely to spread from a vertex to adjacent vertices (friends) than to completely unrelated vertices.)

2 An overview of Ramsey's theorem

2.1 Motivation

One of the reasons to study both clique number *and* independence number is to look at how they are related in the same graph. Intuitively, if a graph G has a low clique number $\omega(G)$, then it doesn't have too many edges—and its independence number $\alpha(G)$ is high. Conversely, if $\alpha(G)$ is low, then G must have a lot of edges to prevent large independent sets from occurring—and this results in a higher value of $\omega(G)$.

How do we quantify this? The answer is Ramsey's theorem. Before stating this theorem in full, let's look at a traditional simple case. This case is often stated informally as "if you have a party with 6 guests, either you can find 3 guests that all know each other, or you can find 3 guests that are all meeting for the first time". More abstractly:

Theorem 2.1. *If a graph G has at least 6 vertices, then either $\alpha(G) \geq 3$ or $\omega(G) \geq 3$.*

Proof. Let v be an arbitrary vertex of G. We consider two cases:

Case 1: $\deg(v) \geq 3$. Then we can choose three vertices w_1, w_2, w_3 all adjacent to v. Even a single edge w_1w_2, w_1w_3, or w_2w_3 between these creates a 3-vertex clique: $\{v, w_1, w_2\}$, $\{v, w_1, w_3\}$, or $\{v, w_2, w_3\}$ respectively. However, if none of these edges exist, then $\{w_1, w_2, w_3\}$ is a 3-vertex independent set.

Case 2: $\deg(v) \leq 2$. In this case, since G has at least 5 vertices other than v, and at most two of them are adjacent to v, we can do the opposite: choose three vertices u_1, u_2, u_3 none of which are adjacent to v. From here, Case 2 is handled similarly to Case 1. If all three edges u_1u_2, u_1u_3, and u_2u_3 are present, then $\{u_1, u_2, u_3\}$ is a clique. So suppose one of these edges, u_iu_j for some $i \neq j$, is absent. Then $\{v, u_i, u_j\}$ is an independent set.

In both cases, we find either a clique or an independent set with 3 vertices, so either $\alpha(G) \geq 3$ or $\omega(G) \geq 3$. □

The number 6 is the best possible in this theorem, since C_5 is a 5-vertex graph with $\alpha(C_5) = \omega(C_5) = 2$.

Ramsey's theorem says, more generally, that the same thing happens for k-vertex cliques or independent sets. For every k, there is a number $R(k)$ (called the k^{th} **Ramsey number**) such that every graph G with $R(k)$ or more vertices has either $\alpha(G) \geq k$ or $\omega(G) \geq k$.

Finding these Ramsey numbers is an open area of study. Today, we will see some upper and lower bounds on $R(k)$.

Ramsey *theory* is an area of study in math that starts here but goes beyond graph theory. It is all about finding some kind of small ordered structure in a large, chaotic world. (In Ramsey's theorem, our "world" is an arbitrary large graph, and our ordered structure is either a clique or an independent set.)

2.2 Upper bounds

Theorem 2.2. *For all k, a graph G on 2^{2k} or more vertices has either $\alpha(G) \geq k$ or $\omega(G) \geq k$. In particular, Ramsey's theorem holds, and $R(k) \leq 2^{2k}$.*

Proof. We are trying to prove that G has either a clique or an independent set—so we want to give an algorithm that will do it. But we can't guarantee specifically a clique or specifically an independent set, so what should we be looking for?

The answer is that we will begin by finding a sequence of vertices $v_1, v_2, \dots, v_{2k}$ such that every vertex v_i is either adjacent to **all** of the vertices $v_{i+1}, \dots, v_{2k}$, or to **none** of them. Call this an "all-or-nothing" sequence. This is a sort of mix between a clique (which chooses the "all" option for every vertex) and an independent set (which chooses the "none" option for every vertex).

We'll build up the all-or-nothing sequence one vertex at a time. We'll also maintain a pool S of "suitable vertices" that can be used to continue the sequence. Initially, the sequence has no elements, and S is all of $V(G)$.

We begin by choosing v_1 to be any element of S: that is, any vertex. (Remove v_1 from S as we do so.) Then, we reduce the pool of available vertices further. If $\deg(v_1) < 2^{2k-1}$, we throw away all vertices adjacent to v_1, ensuring that v_1 is adjacent to **all** the vertices left in S. If $\deg(v_1) \geq 2^{2k-1}$, instead we throw away all vertices not adjacent to v_1, ensuring that v_1 is adjacent to **none** of the vertices left in S. We make the choice this way to ensure that $|S| \geq 2^{2k-1}$ after this step is done.

We continue in the same way. Suppose that we've already chosen $v_1, v_2, \dots, v_{i-1}$. Then we choose v_i to be any of the suitable vertices in S, and remove v_i from S. To continue, we pare down S to either the subset which is adjacent to v_i, or the subset not adjacent to v_i, whichever is larger. When we make the choice in this way, $|S|$ decreases at most by a factor of 2.

The set S keeps shrinking, and after $|S| = 0$, of course we can't choose any more vertices. But because $|S|$ is at worst cut in half after every step, we still have $|S| \geq 2^{2k-i}$ when i vertices have been chosen. In particular, $|S|$ will stay positive long enough to choose an all-or-nothing sequence of length $2k$.

Once we have this sequence what do we do? Well, with $2k$ vertices to choose from, either k of them choose the "all" option (and are adjacent to all following vertices) or k of them choose the "none" option (and are adjacent to none of the following vertices). In the first case, we see a k-vertex clique; in the second case, we see a k-vertex independent set. $\square$

This proof is not written in a careful way that tries to get the best bound; in particular, it says $R(3) \leq 2^6 = 64$, and we already know that $R(3) = 6$. We can improve the bound in some ways (see the practice problems).

However, even our best upper bounds on $R(k)$ do not grow significantly slower than 2^{2k}. In 2023, a breakthrough was made: the upper bound was improved to roughly $R(k) \leq 3.999^k$. Even this much is a big deal!

2.3 Lower bounds

So is $R(k)$ as big as that? The answer is both yes and no (and also maybe).

If you look for lower bounds on $R(k)$, you will probably not find very good ones. It turns out to be very difficult to construct graphs by hand that have a small clique number as well as a small independence number.

However, such graphs do exist. And even though they're hard for us to find, they're not rare: we can find one just by picking it at random.

Theorem 2.3. *For $k \geq 3$, let G be a graph with $n = 2^{k/2-1}$ vertices; flip a coin for every possible edge to see if it's present or absent. Then there is a positive probability that $\alpha(G) < k$ and $\omega(G) < k$.*

(Note: when k is odd, $2^{k/2-1}$ is not an integer, but we can round down; this does not hurt the proof.)

Proof. For any given set of k vertices, what is the probability that they form a clique? Well, there are $\binom{k}{2} = \frac{k(k-1)}{2}$ edges between them. All of the coinflips for those edges have to go one way, for a probability of $(\frac{1}{2})^{k(k-1)/2}$.

That's not the whole story, though, because there are many k-vertex sets that could form a clique. As an upper bound, there are at most $n^k = 2^{k(k/2-1)}$. This is an overcount—n^k counts an ordered sample with replacement, and here we don't need the order and don't want to take vertices with replacement. But for the upper bound, an overcount is fine.

If you have N events and each happens with probability p, then the probability that any of them happen is at most Np: this worst case is achieved when the events are disjoint. If they are not disjoint, we have to subtract the overlaps, so Np is only an upper bound; again, that's fine here. We have at most $2^{k(k/2-1)}$ events and each happens with probability $(\frac{1}{2})^{k(k-1)/2}$, so the probability that any of them happen is at most $2^{k(k/2-1)} \cdot 2^{-k(k-1)/2} = 2^{-k/2}$.

This is our upper bound for the probability that G has a k-vertex clique. We get a similar upper bound for the probability that G has a k-vertex independent set. So the overall probability that $\alpha(G) \geq k$ or $\omega(G) \geq k$ is at most $2^{-k/2} + 2^{-k/2} = 2^{1-k/2}$. This probability gets very very small for large k, but even for $k = 3$ it is $2^{-1/2} = \frac{\sqrt{2}}{2} \approx 0.707$. That's less than 1. $\square$

What we see here is a use of the probabilistic method: rather than find an example explicitly, we show that if we pick a random example, it has a positive (and often quite high) chance of working. In particular, graphs like this must exist: if they did not exist, the probability would be 0. We conclude that $R(k) > 2^{k/2-1}$.

3 Practice problems

1. Consider the graph with ab vertices corresponding to pairs (x, y) where $1 \le x \le a$ and $1 \le y \le b$, and where two vertices (x, y) and (x', y') are adjacent whenever $x \ne x'$. (This is a complete a-partite graph where each part has size b.)

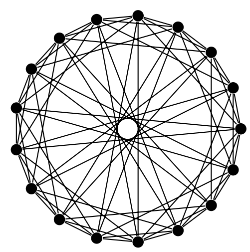
   (a) What is the clique number of this graph? Describe what the largest cliques look like.

   (b) What is the independence number of this graph? Describe what the largest independent sets look like.

2. If we really really just want independent sets and not cliques, we can modify the algorithm in the proof of Theorem 2.2 to always replace S by the subset of S not adjacent to v_i. Of course, now it's possible for $|S|$ to decrease much faster in some graphs.

   Prove that in a 100-vertex graph with maximum degree 10, the algorithm finds an independent set of size 10. How well does it do in general, in an n-vertex graph G with maximum degree $\Delta(G)$?

3. The graph below is defined by a simple rule: the 17 vertices are spaced equally around a circle, and there is an edge between two vertices when they are 1, 2, 4, or 8 steps away.

   (a) What is the independence number of this graph?

   (b) What is the clique number of this graph?

   (c) This graph is famous for being an example that proves a lower bound for a Ramsey number $R(k)$. What lower bound does it prove, for which value of k?

4. By being more careful in the proof of Theorem 2.2, we can improve the upper bound it gives. See if you can get the proof to give 2^{2k-2} as an upper bound instead.

5. The reason Theorem 2.3 uses a randomly chosen graph is that non-random graphs that do equally well are hard to describe. This problem has one possible construction that doesn't do nearly as well as the random graph, but does better than many other possibilities.

   Let G_1 be the 5-cycle C_5. Then, for each $k > 1$, construct G_k by starting with G_{k-1}, and then replacing

   - Each vertex v of G_{k-1} by five vertices v_1, v_2, v_3, v_4, v_5 with a 5-cycle through them;
   - Each edge vw of G_{k-1} by 25 edges $v_i w_j$ for $1 \le i, j \le 5$.

   In other words, we replace the vertices of G_{k-1} by copies of C_5.

   (a) Prove that $\alpha(G_2) = \omega(G_2) = 4$. What do the cliques and independent sets look like?

(b) Prove that $\alpha(G_k) = \omega(G_k) = 2^k$. (Induct on k.)

(c) We get a lower bound $R(17) > n$ by finding an n-vertex graph G with $\alpha(G) \leq 16$ and $\omega(G) \leq 16$.

What lower bound does the construction in this problem give for $R(17)$? How does it compare to the lower bound from Theorem 2.3?

What if we try to bound $R(33)$ instead?