Graph Theory

Colorings, cliques, independent and dominating sets

Instructor: *Nicolò Cesa-Bianchi* version of March 12, 2024

When we model a certain problem instance in terms of a graph, we may be interested in defining specific graph-theoretical properties that lead to further insights for solving the starting problem. Some of these properties can be quantified in terms of real numbers. They are typically referred to as graph parameters. A **graph parameter** is a function $\phi\colon \mathcal{G} \to \mathbb{R}$ that maps the set of simple graphs $\mathcal{G}$ to real values.

Graph parameters are called graph invariants because these properties only depend on the structure of the graphs, and thus are invariant under graph isomorphisms. Two graphs $G = (V, E)$ and $H = (V', E')$ are isomorphic if there exists a bijective function $f\colon V \to V'$ such that $(u, v) \in E$ if and only if $\big(f(u), f(v)\big) \in E'$. In this case, we write $G \simeq H$ and the function f is called **graph isomorphism**. Assuming G and H are isomorphic, we know that $\phi(G) = \phi(H)$ for any graph invariant ϕ. We now introduce some of the most common graph parameters.

The first graph parameter we introduce is based on the notion of independent set. A subset $U \subseteq V$ is independent in $G = (V, E)$ if no two vertices in U are neighbors. We can equivalently define a set U to be independent when the induced subgraph $G[U]$ has no edges, that is, $G[U] = (U, \emptyset)$. The **independence number** $\alpha(G)$ is the size of the largest independent set in G. The associated decision problem is NP-complete. It is easy to prove the sum of the independence number and the size of the smallest vertex cover[1] is always equal to the order of the graph.

The definition of independent sets is connected to that of cliques. The **clique number** $\omega(G)$ is the size of the largest clique in G. The associated decision problem is NP-complete (equivalent to independent set in the complement graph $\overline{G}$ of G).

A third well-known graph-theoretical construct that we focus on is that of vertex colorings. A vertex coloring $c : V \to \{1, \dots, k\}$ of $G = (V, E)$ is an assignment of colors $\{1, \dots, k\}$ to vertices such that $(x, y) \in E$ implies $c(x) \neq c(y)$. G is k-colorable if there exists a coloring c with codomain of size k. Equivalently, G is k-colorable if there exists a partition $V_1, \dots, V_k$ of V such that each subset V_i is independent in G. The **chromatic number** $\chi(G)$ is the smallest integer k such that G is k-colorable. The associated decision problem is NP-complete except for $k \leq 2$. Here are some easy facts.

- Any bipartite graph is 2-colorable (equivalent to having no odd-length cycle in the graph).
- For the clique, $\chi(K_n) = n$. This implies $\omega(G) \leq \chi(G)$ for all G.
- For the cycle, $\chi(C_{2n}) = 2$ and $\chi(C_{2n+1}) = 3$.

Another well-known fact is given by the four-color theorem, which states that $\chi(G) \leq 4$ for all planar[2] graphs G. More generally, we clearly have that $1 \leq \chi(G) \leq n$, but we can further prove that $\chi(G) \leq 1 + \sqrt{2E}$ as shown next.

[1] A subset $U \subseteq V$ is a vertex cover of $G = (V, E)$ if all edges in E have at least an endpoint in U.

[2] A graph is planar when it can be drawn so that if any two edges intersect, they only do so in a vertex.

Fact 1 *For all G,*

$$\chi(G) \leq \frac{1}{2} + \sqrt{2|E| + \frac{1}{4}}\ .$$

PROOF. Let $k = \chi(G)$ and $V_1, \ldots, V_k$ be the partition of V induced by the k-coloring. Due to the minimality of k, there is at least one edge in E for any pair (V_i, V_j) with $i \neq j$. Therefore, $|E| \geq \binom{k}{2}$. Solving for k gives the result. □

The chromatic number cannot be large when all vertices have small degrees. We can then prove another upper bound on the chromatic number, which is known to be tight when $G = K_n$ or $G = C_{2n+1}$.

Fact 2 *For all G, $\chi(G) \leq \Delta(G) + 1$.*

PROOF. If $\chi(G) > k$, then there exists $v \in V$ with $d(v) \geq k$. Hence, if $\Delta(v) < k$ then $\chi(G) \leq k$, which implies $\chi(G) \leq \Delta(v) + 1$. □

Given that each color uniquely corresponds to an independent set, a small independence number implies a large chromatic number.

Fact 3 *For all G, $\chi(G)\alpha(G) \geq |V|$.*

PROOF. Let $k = \chi(G)$ and $V_1, \ldots, V_k$ the partition of V induced by a k-coloring of G. Since each V_i is independent in G,

$$|V| = \sum_{i=1}^{k} |V_i| \leq \sum_{i=1}^{k} \alpha(G) \leq \chi(G)\alpha(G)$$

concluding the proof. □

The next important result shows that a small average degree implies a large independence number. We being by introducing a fundamental tool that will be useful within the proof of this result.

Fact 4 (Jensen's inequality) *Let $X \in \mathbb{R}$ be a real-valued random variable such that $X \sim P$ for some distribution P. Let $f \colon \mathbb{R} \to \mathbb{R}$ be a convex function. Then, $f(\mathbb{E}_P[X]) \leq \mathbb{E}_P[f(X)]$.*

In particular, we can consider the distribution P to be the empirical distribution with probability mass function $P(x) = \sum_{i=1}^{m} \frac{\mathbb{I}\{x=x_i\}}{m}$ over a certain set $\{x_1, \ldots, x_m\}$. Given a function $h \colon \mathbb{R} \to \mathbb{R}$, this implies that the expectation computed over P is $\mathbb{E}_P[h(X)] = \frac{1}{m}\sum_{i=1}^{m} h(x_i)$. When h is the identity function the expectation above becomes the sample mean and Jensen's inequality states that

$$f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) \leq \frac{1}{m}\sum_{i=1}^{m} f(x_i)$$

Additionally, we introduce the notation $C(v) = N(v) \cup \{v\}$ for the neighborhood of v that includes v itself. Now we have all the tools required for proving the following theorem.

Theorem 5 (Turán, 1941) *For all G, $\alpha(G)\big(d(G) + 1\big) \geq |V|$.*

Algorithm 1 Greedy construction of an independent set

Input: graph $G = (V, E)$
$i \leftarrow 1$; $G_1 \leftarrow G$
while G_i is not empty **do**
  $v_i \in \operatorname{argmin}_{v \in V_i} d_i(v)$
  $G_{i+1} \leftarrow G_i - C_i(v_i)$ ▷ remove $C_i(v_i)$ and all incident edges from G_i
  $i \leftarrow i + 1$
end while
return $\{v_1, \ldots, v_{i-1}\}$

PROOF. Consider the following greedy algorithm to construct an independent set: pick the vertex of smallest degree and remove it from the graph together with its neighborhood. Iterate on the remaining graph until there are no more vertices to pick.

Let $\{v_1, \ldots, v_s\}$ be the set of s vertices picked by the algorithm. Furthermore, let $G_1, \ldots, G_{s+1}$ be the sequence of graphs generated by the algorithm: each $G_i = (V_i, E_i)$ has degree function $d_i = d_{G_i}$ and "extended" neighborhood function $C_i(v) = N_i(v) \cup \{v\}$. A graph G_{i+1} is obtained from G_i by removing the vertices in $C_i(v_i)$ along with the edges that have an endpoint in $C_i(v_i)$.

Since the algorithm removes at least one vertex (v_i) at every step i, the algorithm always terminates in a finite number $s \leq n$ of iterations. Notice that we can find a tighter upper bound on the number of iterations. Consider the vertices $v_1, \ldots, v_s$ picked by the algorithm. Clearly, this set of vertices forms an independent set of size $s \leq \alpha(G)$. This follows from the fact that, when the algorithm removes a vertex v_i, it removes all its neighbors $N_i(v_i)$ along with it. Therefore, it cannot be the neighbor of any vertex picked either before or after it.

For the purpose of this proof, introduce the quantity

$$Q(G') = \sum_{u \in V(G')} \frac{1}{1 + d(u)} \ ,$$

where $d(u)$ is the degree of u in the *initial* graph G. At each step i, this quantity goes down by

$$Q(G_i) - Q(G_{i+1}) = \sum_{u \in V_i} \frac{1}{1 + d(u)} - \sum_{u \in V_{i+1}} \frac{1}{1 + d(u)} = \sum_{u \in C_i(v_i)} \frac{1}{1 + d(u)} \ .$$

Considering the fact that the degree of a vertex can only decrease until its removal from the graph, we know that $d(u) \geq d_i(u)$ for all $u \in V_i$. Thus, the variation of Q at step i is bounded by

$$\sum_{u \in C_i(v_i)} \frac{1}{1 + d(u)} \leq \sum_{u \in C_i(v_i)} \frac{1}{1 + d_i(u)} \leq \sum_{u \in C_i(v_i)} \frac{1}{1 + d_i(v_i)} = \frac{|C_i(v_i)|}{1 + d_i(v_i)} = 1 \ ,$$

where we also use the fact that $d_i(u) \geq d_i(v_i)$ for any $u \in V_i$ by choice of v_i. Since it takes $s \leq \alpha(G)$ steps until $Q(G)$ goes to zero (indeed, $Q(G_{s+1}) = 0$ because G_{s+1} is empty), and we decrease $Q(G)$ by at most one in each step, then

$$Q(G) = Q(G_1) - Q(G_{s+1}) = \sum_{i=1}^{s} \underbrace{\big(Q(G_i) - Q(G_{i+1})\big)}_{\leq 1} \leq s \leq \alpha(G) \ . \tag{1}$$

Now we just observe that

$$\frac{Q(G)}{|V|} = \frac{1}{|V|}\sum_{v\in V}\frac{1}{1+d(v)} \geq \frac{1}{1+\frac{1}{|V|}\sum_{v\in V} d(v)} = \frac{1}{1+d(G)} \tag{2}$$

because of Jensen's inequality applied to the function $f(x) = \frac{1}{1+x}$, which is convex for $x > -1$. We conclude the proof by putting together Equations (1) and (2). $\square$

Turán's theorem also shows that a large average degree implies a large clique number.

Corollary 6 *For all G, $\omega(G)\big(|V| - d(G)\big) \geq |V|$.*

Proof. Let $n = |V|$ and $\overline{G} = (V, \overline{E})$ be the complement of G, where $e \in \overline{E}$ if and only if $e \notin E$. If $d_G(v)$ is the degree of v in G, then its degree in $\overline{G}$ is $d_{\overline{G}}(v) = n - 1 - d_G(v)$. Consequently, if $d(G)$ is the average degree in G and $d(\overline{G})$ is the one in $\overline{G}$, then

$$d(\overline{G}) = \frac{1}{n}\sum_{v\in V} d_{\overline{G}}(v) = \frac{1}{n}\sum_{v\in V}\big(n - 1 - d_G(v)\big) = n - 1 - d(G) \ .$$

As an independent set in $\overline{G}$ corresponds to a clique in G, we can apply Turán's theorem and obtain

$$\omega(G) = \alpha(\overline{G}) \geq \frac{n}{d(\overline{G}) + 1} = \frac{n}{n - d(G)} \ ,$$

concluding the proof. $\square$

In this last part, we focus on a new graph invariant based on the definition of dominating set. A subset $U \subseteq V$ is dominating in $G = (V, E)$ if every vertex in $V \setminus U$ has a neighbor in U. The **domination number** $\gamma(G)$ is the size of the smallest dominating set in G. The associated decision problem is NP-complete.

It is fairly easy to show that the domination number is always smaller than the independence number.

Fact 7 *For all G, $\gamma(G) \leq \alpha(G)$.*

Proof. Let U be an independent set of maximum cardinality. Then U is a dominating set. Indeed, if U is not dominating then there is $x \in V \setminus U$ without neighbors in U. But then $U \cup \{x\}$ would be an independent set larger than U, leading to a contradiction. $\square$

The next result shows that if all vertices have a large degree, then the domination number must be small.

Theorem 8 (Arnautov, 1974; Payan, 1975) *For all G,*

$$\gamma(G)\frac{1+\delta(G)}{1+\ln\big(1+\delta(G)\big)} \leq |V| \ .$$

Proof. Let $n = |V|$ and $\delta = \delta(G)$. We run a greedy algorithm choosing the vertices for the dominating set one by one, where in each step a vertex that covers the maximum number of yet uncovered vertices is picked (as long as there are any), where an uncovered vertex does not lie in the union of the sets $C(v)$ of the vertices v chosen by the algorithm so far. Let $\mathcal{U}$ be the set of currently uncovered vertices and let $r = |\mathcal{U}|$. Then, the next vertex v' chosen by the algorithm satisfies

$$\begin{aligned}
\sum_{u\in\mathcal{U}} \mathbb{I}\{u \in C(v')\} &= \max_{v\in V} \sum_{u\in\mathcal{U}} \mathbb{I}\{u \in C(v)\} \\
&\geq \mathbb{E}\left[\sum_{u\in\mathcal{U}} \mathbb{I}\{u \in C(X)\}\right] \qquad\qquad (X \in V \text{ chosen uniformly at random}) \\
&= \sum_{u\in\mathcal{U}} \mathbb{P}\big(u \in C(X)\big) \\
&= \sum_{u\in\mathcal{U}} \frac{|C(u)|}{n} \\
&\geq \sum_{u\in\mathcal{U}} \frac{\delta+1}{n} = \frac{r(\delta+1)}{n} \quad .
\end{aligned}$$

Adding this v' to the set of chosen vertices, we observe that the number of uncovered vertices is now at most $r(1-(\delta+1)/n)$. It follows that in each iteration of the above procedure the number of uncovered vertices decreases by a factor of $1-(\delta+1)/n$. The number m of steps it takes so that $|\mathcal{U}| \leq \frac{n}{\delta+1}$ satisfies $n\big(1-\frac{\delta+1}{n}\big)^m \leq \frac{n}{\delta+1}$. Using $1-x \leq e^{-x}$, this implies

$$n \cdot \exp\left(-\frac{\delta+1}{n} m\right) \leq \frac{n}{\delta+1} \ ,$$

which solved for m gives $m \geq \frac{n}{\delta+1}\ln(\delta+1)$. Then, we can stop the greedy procedure as soon as $s = |\mathcal{U}| \leq n/(\delta+1)$. Adding this set of $s \leq n/(\delta+1)$ yet uncovered vertices gives the desired result (the procedure returns a dominating set of size $m + s \geq \gamma(G)$). $\square$