Chapter 2

Adjacency matrix and its rank

Perhaps the easiest way of assigning a vector to each node is to use the corresponding column of the adjacency matrix. Even this easy construction has some applications, and we will discuss one of them.

2.1 Neighborhoods, rank and size

If we know the rank of the adjacency matrix of a graph, how large can the graph be? Obviously we have to make some assumption, since we can replace any node by an arbitrary number of twins without increasing the rank of the adjacency matrix. However, if we exclude twins, then the question becomes meaningful.

An almost trivial answer is given by the following argument. We call the rank of the adjacency matrix A_G of graph G simply the *rank of* G, and denote it by $r = \mathrm{rk}(G)$. Let, say, the first r columns of the adjacency matrix form a basis of the column space. Two rows that agree in their first r positions will agree everywhere, which is impossible if there are no twins. So the initial r-tuples of rows will be all different, and hence there are at most 2^r rows.

This argument works over any field, and no better bound can be given over, say, $GF(2)$ (see Exercise 2.1). However, over the real field we can use geometry, not just linear algebra, and prove a substantially better (almost optimal) bound.

Theorem 2.1.1 *Let G be a twin-free graph on n nodes, and let $r = \mathrm{rk}(A_G)$. Then $n = O(2^{r/2})$.*

Before turning to the proof, we recall a result from discrete geometry. The *kissing number* $s(d)$ in $\mathbb{R}^d$ is the largest integer N such that there are N non-intersecting unit balls touching a given unit ball. The exact value of $s(d)$ is only known for special dimensions d, but the following bound, which follows from tighter estimates, will be enough for us:

$$s(d) = O(2^{d/2}). \tag{2.1}$$

For technical reasons, we state this fact as follows. There is a constant $C > 16$ such that, setting $f(r) = C2^{r/2} - 16$, every set of more than $f(d)$ vectors in $\mathbb{R}^d$ contains two vectors such that angle between them is less than $\pi/3$.

Lemma 2.1.2 *If a graph G of rank r has $n > s(r+1)$ nodes, than it has two nodes i and j such that $|N(i)\triangle N(j)| < n/4$.*

Proof. Let $\mathbf{a}_i$ denote the column of A_G corresponding to $i \in V$, and let $\mathbf{u}_i = \mathbb{1} - 2\mathbf{a}_i$. Clearly the vectors $\mathbf{u}_i$ are ± 1-vectors, which belong to an (at most) $(r+1)$-dimensional subspace. Applying the kissing number bound to the vectors $\mathbf{u}_i$, we get that there there are two vectors $\mathbf{u}_i$ and $\mathbf{u}_j$ forming an angle less than $\pi/3$. For two ± 1 vectors, this means that $\mathbf{u}_i$ and $\mathbf{u}_j$ differ in fewer than $n/4$ positions. The vectors $\mathbf{a}_i$ and $\mathbf{a}_j$ also differ in these positions only. □

We also need a simple facts about twins in induced subgraphs. If we delete a node from a graph G that has a twin, then no new twins are created (if pair i, j of remaining nodes is distinguished by some node of G, and if this gets deleted, then its twin remains and still distinguishes i and j). Deleting a node that has a twin does not change the rank of the graph.

Lemma 2.1.3 *If two nodes $i, j \in V$ of a graph G are not twins, but they are twins in an induced subgraph H, then $\mathrm{rk}(H) \le \mathrm{rk}(G) - 2$.*

Proof. Indeed, A_G must contain a column that distinguishes rows i and j, and so it is not in the linear span of the columns in X. Adding this column to A_H increases its rank. Adding the corresponding row increases the rank further. □

Proof of Theorem 2.1.1. We prove that $n \le f(r)$ for every graph on n nodes and of rank r, by induction on r. For $r = 1, 2, 3$ this is easily checked. So we may suppose that $r > 3$.

If $|N(i)\triangle N(j)| \ge n/4$ for any two nodes, then Lemma 2.1.2 implies that $n \le s(r+1) < f(r)$. So we may assume that there are two nodes i and j such that $|N(i)\triangle N(j)| < n/4$. The set $X = V \setminus (N(i)\triangle N(j))$ induces a subgraph in which i and j are twins, and hence by Lemma 2.1.3, $\mathrm{rk}(G[X]) \le r - 2$.

Let $G[Z]$ be a largest induced subgraph of G with $\mathrm{rk}(G[Z]) \le r - 1$. Since X induces one of these subgraphs, we have

$$|Z| \ge |X| > \frac{3}{4}n. \tag{2.2}$$

If $G[Z]$ does not have twins, then by the induction hypothesis applied to $G[Z]$,

$$n < \frac{4}{3}|Z| \le \frac{4}{3}f(r-1) < f(r).$$

So we may assume that $G[Z]$ has twins. By Lemma 2.1.3, we have $\mathrm{rk}(G[Z]) \le r - 2$.

Let $u \in V \setminus Z$. By the maximality of Z, the rows of A_G corresponding to $Z \cup \{u\}$ span its row-space. This implies that u must distinguish every twin pair in Z, and so u is connected to exactly one member of every such pair. This in particular implies that $G[Z]$ does not have three mutually twin nodes. Let T be the set of nodes that are twins in $G[Z]$ and are adjacent to u, and let T' be the set of twins of nodes of T. Let $U = Z \setminus (T \cup T')$.

Next consider any other $v \in V \setminus Z$. We show that it is either connected to all of nodes in T and none in T', or the other way around. Row v of A_G is a linear combination of rows corresponding to Z and u; let, say, the coefficient of row u in this linear combination be positive. Then for every twin pair $\{i, j\}$ (where $i \in T$ and $j \in T'$) we have $a_{u,i} > a_{u,j}$, but (by the definition of twins) $a_{w,i} = a_{w,j}$ for all $w \in Z$, and hence $a_{v,i} > a_{v,j}$. This means that v is adjacent to i, but not to j, and so we get that v is adjacent to all nodes of T but not to any node of T'. Thus we have a decomposition $V \setminus Z = Y \cup Y'$, where every node of Y is connected to all nodes of T but to no node of T', and for nodes in Y', the other way around.

So G has the structure in Figure 2.1.

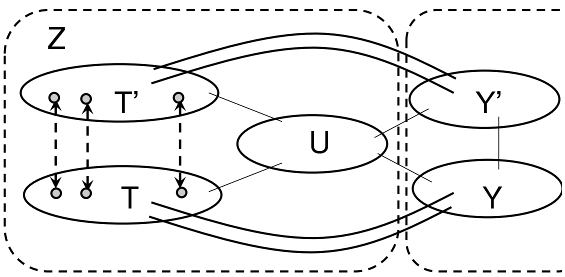

Figure 2.1: Node sets in the proof of Theorem 2.1.1. Dashed arrows indicate twins in $G[Z]$.

The graph $G[T \cup U]$ is obtained from $G[Z]$ by deleting one member of each twin-pair, hence it is twin-free. Since $\mathrm{rk}(G[T \cup U]) = \mathrm{rk}(G[Z]) \le r - 2$, we can apply the induction hypothesis to get

$$|T| + |U| \le f(r - 2). \tag{2.3}$$

If $|Y \cup Y'| \le 16$, then

$$n = 2|T| + |U| + |Y \cup Y'| \le 2f(r - 2) + 16 \le f(r),$$

and we are done. So we may assume that (say) $|Y'| > 8$. Since the rows in Y' are all different, this implies that they form a submatrix of rank at least 4. Using that in their columns corresponding to T these rows have zeros, we see that the matrix formed by the rows in $Y' \cup T$ has rank at least $\mathrm{rk}(G[T]) + 4$. This implies that $\mathrm{rk}(G[T]) \le r - 4$.

If $G[T]$ is not twin-free, then any twins of it remain twins in $G - U$: no node in T' can distinguish them (because its twin does not), and no node in $V \setminus Z$ can distinguish them, as we have seen. So $G - U$ is not twin-free, and hence $\mathrm{rk}(G-U) \le r-2$ by Lemma 2.1.3. By the maximality of Z, we have $|V \setminus U| \le |Z|$. Using (2.3), this gives

$$n \le |Z| + |U| = 2|T| + 2|U| \le 2f(r-2) < f(r).$$

On the other hand, if $G[T]$ is twin-free, then we can apply induction:

$$|T| \le f(r-4). \tag{2.4}$$

Using (2.2), (2.3) and (2.4), we get that

$$n < \frac{4}{3}|Z| = \frac{4}{3}(2|T| + |U|) \le \frac{4}{3}(f(r-2) + f(r-4)) < f(r).$$

□

The bound in Theorem 2.1.1 is sharp up to the constant. (This shows that even though the bound on the kissing number could be improved, this would not give an improvement of the whole argument.) Let us define a sequence of graphs G_r recursively: Let $G_1 = K_2$, and given G_r ($r \ge 2$), let G_{r+1} be obtained from G_r by doubling each node (creating $|V(G_r)|$ pairs of twins), adding a new node, and connecting it to one member of each twin pair. It is easy to see by induction that G_r is twin-free, $|V(G_r)| = 3 \cdot 2^{r-1} - 1$, and $\mathrm{rk}(A_{G_r}) = 2r$.

As a corollary, we get that if the rank of the adjacency matrix of a simple graph (which may have twins) is r, then its chromatic number is at most $f(r)$ (since twins can be colored with the same color). It is not known how large the chromatic number can be in terms of r.

Exercise 2.1 Show that for every even $r \ge 2$, there is a twin-free simple graph on 2^r nodes whose adjacency matrix has rank r over $GF(2)$. Find the best construction for odd r.

Exercise 2.2 Prove that every graph on n nodes contains two nodes i and j with $|N(i) \triangle N(j)| < n/2$. Show that this bound is best possible.