Given a family of sets $A_1, \ldots, A_N$, their *average size* is

$$\frac{1}{N}\sum_{i=1}^{N}|A_i|.$$

The following lemma says that, if the average size of sets is large, then some two of them must share many elements.

Lemma 2.2. *Let X be a set of n elements, and let $A_1, \ldots, A_N$ be subsets of X of average size at least n/w. If $N \geq 2w^2$, then there exist $i \neq j$ such that*

$$|A_i \cap A_j| \geq \frac{n}{2w^2}. \tag{2.3}$$

Proof. Again, let us just count. On the one hand, using Jensen's inequality (1.15) and equality (1.10), we obtain that

$$\sum_{x\in X} d(x)^2 \geq \frac{1}{n}\Big(\sum_{x\in X} d(x)\Big)^2 = \frac{1}{n}\Big(\sum_{i=1}^{N}|A_i|\Big)^2 \geq \frac{nN^2}{w^2}.$$

On the other hand, assuming that (2.3) is false and using (1.11) and (1.12) we would obtain

$$\begin{aligned}\sum_{x\in X} d(x)^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N}|A_i\cap A_j| = \sum_{i}|A_i| + \sum_{i\neq j}|A_i\cap A_j| \\ &< nN + \frac{nN(N-1)}{2w^2} = \frac{nN^2}{2w^2}\Big(1+\frac{2w^2}{N}-\frac{1}{N}\Big) \leq \frac{nN^2}{w^2},\end{aligned}$$

a contradiction. □

Lemma 2.2 is a very special (but still illustrative) case of the following more general result.

Lemma 2.3 (Erdős 1964b)**.** *Let X be a set of n elements $x_1, \ldots, x_n$, and let $A_1, \ldots, A_N$ be N subsets of X of average size at least n/w. If $N \geq 2kw^k$, then there exist $A_{i_1}, \ldots, A_{i_k}$ such that $|A_{i_1}\cap\cdots\cap A_{i_k}| \geq n/(2w^k)$.*

The proof is a generalization of the one above and we leave it as an exercise (see Exercises 2.8 and 2.9).

2.2 Graphs with no 4-cycles

Let H be a fixed graph. A graph is H-free if it does not contain H as a subgraph. (Recall that a *subgraph* is obtained by deleting edges and vertices.) A typical question in graph theory is the following one:

How many edges can a H-free graph with n vertices have?

That is, one is interested in the maximum number $\mathrm{ex}(n,H)$ of edges in a H-free graph on n vertices. The graph H itself is then called a "forbidden subgraph."

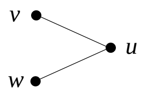
Let us consider the case when forbidden subgraphs are *cycles*. Recall that a cycle C_k of length k (or a k-cycle) is a sequence $v_0, v_1, \ldots, v_k$ such that $v_k = v_0$ and each subsequent pair v_i and v_{i+1} is joined by an edge.

If $H = C_3$, a triangle, then $\mathrm{ex}(n, C_3) \geq n^2/4$ for every even $n \geq 2$: a complete bipartite $r \times r$ graph $K_{r,r}$ with $r = n/2$ has no triangles but has $r^2 = n^2/4$ edges. We will show later that this is already optimal: any n-vertex graph with more than $n^2/4$ edges must contain a triangle (see Theorem 4.7). Interestingly, $\mathrm{ex}(n, C_4)$ is much smaller, smaller than $n^{3/2}$.

Theorem 2.4 (Reiman 1958)**.** *If $G = (V,E)$ on n vertices has no 4-cycles, then*

$$|E| \leq \frac{n}{4}(1 + \sqrt{4n-3}) .$$

Proof. Let $G = (V,E)$ be a C_4-free graph with vertex-set $V = \{1, \ldots, n\}$, and $d_1, d_2, \ldots, d_n$ be the degrees of its vertices. We now count in two ways the number of elements in the following set S. The set S consists of all (ordered) pairs $(u, \{v,w\})$ such that $v \neq w$ and u is adjacent to both v and w in G. That is, we count all occurrences of "cherries"



in G. For each vertex u, we have $\binom{d_u}{2}$ possibilities to choose a 2-element subset of its d_u neighbors. Thus, summing over u, we find $|S| = \sum_{u=1}^{n} \binom{d_u}{2}$. On the other hand, the C_4-freeness of G implies that no pair of vertices $v \neq w$ can have more than one common neighbor. Thus, summing over all pairs we obtain that $|S| \leq \binom{n}{2}$. Altogether this gives

$$\sum_{i=1}^{n} \binom{d_i}{2} \leq \binom{n}{2}$$

or

$$\sum_{i=1}^{n} d_i^2 \leq n(n-1) + \sum_{i=1}^{n} d_i \, . \tag{2.4}$$

Now, we use the Cauchy–Schwarz inequality

$$\Big(\sum_{i=1}^{n} x_i y_i\Big)^2 \leq \Big(\sum_{i=1}^{n} x_i^2\Big)\Big(\sum_{i=1}^{n} y_i^2\Big)$$

with $x_i = d_i$ and $y_i = 1$, and obtain

$$\left(\sum_{i=1}^{n} d_i\right)^2 \le n \sum_{i=1}^{n} d_i^2$$

and hence by (2.4)

$$\left(\sum_{i=1}^{n} d_i\right)^2 \le n^2(n-1) + n\sum_{i=1}^{n} d_i\,.$$

Euler's theorem gives $\sum_{i=1}^{n} d_i = 2|E|$. Invoking this fact, we obtain

$$4|E|^2 \le n^2(n-1) + 2n|E|$$

or

$$|E|^2 - \frac{n}{2}|E| - \frac{n^2(n-1)}{4} \le 0\,.$$

Solving the corresponding quadratic equation yields the desired upper bound on $|E|$. $\square$

Example 2.5 (Construction of dense C_4-free graphs). The following construction shows that the bound of Theorem 2.4 is optimal up to a constant factor.

Let p be a prime number and take $V = (\mathbb{Z}_p \setminus \{0\}) \times \mathbb{Z}_p$, that is, vertices are pairs (a, b) of elements of a finite field with $a \neq 0$. We define a graph G on these vertices, where (a, b) and (c, d) are joined by an edge iff $ac = b + d$ (all operations modulo p). For each vertex (a, b), there are $p - 1$ solutions of the equation $ax = b + y$: pick any $x \in \mathbb{Z}_p \setminus \{0\}$, and y is uniquely determined. Thus, G is a $(p-1)$-regular graph on $n = p(p-1)$ vertices (some edges are loops). The number of edges in it is $n(p-1)/2 = \Omega(n^{3/2})$.

To verify that the graph is C_4-free, take any two its vertices (a, b) and (c, d). The unique solution (x, y) of the system

$$\begin{cases} ax = b + y \\ cx = d + y \end{cases} \quad \text{is given by} \quad \begin{array}{c} x = (b-d)(a-c)^{-1} \\ 2y = x(a+c) - b - d \end{array}$$

which is only defined when $a \neq c$, and has $x \neq 0$ only when $b \neq d$. Hence, if $a \neq c$ and $b \neq d$, then the vertices (a, b) and (c, d) have precisely one common neighbor, and have no common neighbors at all, if $a = c$ or $b = d$.

2.3 Graphs with no induced 4-cycles

Recall that an *induced subgraph* is obtained by deleting vertices together with all the edges incident to them (see Fig. 2.1).

Theorem 2.4 says that a graph cannot have many edges, unless it contains C_4 as a (not necessarily induced) subgraph. But what about graphs that

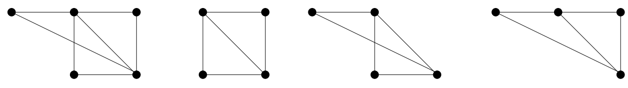


Fig. 2.1 Graph G contains several copies of C_4 as a subgraph, but none of them as an *induced* subgraph.

do not contain C_4 as an *induced* subgraph? Let us call such graphs *weakly C_4-free*.

Note that such graphs can already have many more edges. In particular, the complete graph K_n is weakly C_4-free: in any 4-cycle there are edges in K_n between non-neighboring vertices of C_4. Interestingly, any(!) dense enough weakly C_4-free graph must contain large complete subgraphs.

Let $\omega(G)$ denote the maximum number of vertices in a complete subgraph of G. In particular, $\omega(G) \leq 3$ for every C_4-free graph. In contrast, for *weakly* C_4-free graphs we have the following result, due to Gyárfás, Hubenko and Solymosi (2002).

Theorem 2.6. *If an n-vertex graph $G = (V, E)$ is weakly C_4-free, then*

$$\omega(G) \geq 0.4\frac{|E|^2}{n^3}.$$

The proof of Theorem 2.6 is based on a simple fact, relating the average degree with the minimum degree, as well as on two facts concerning independent sets in weakly C_4-free graphs.

For a graph $G = (V, E)$, let $e(G) = |E|$ denote the number of its edges, $d_{\min}(G)$ the smallest degree of its vertices, and $d_{\text{ave}}(G) = 2e(G)/|V|$ the average degree. Note that, by Euler's theorem, $d_{\text{ave}}(G)$ is indeed the sum of all degrees divided by the total number of vertices.

Proposition 2.7. *Every graph G has an induced subgraph H with*

$$d_{\text{ave}}(H) \geq d_{\text{ave}}(G) \qquad \textit{and} \qquad d_{\min}(H) \geq \frac{1}{2}\, d_{\text{ave}}(G).$$

Proof. We remove vertices one-by-one. To avoid the danger of ending up with the empty graph, let us remove a vertex $v \in V$ if this does not decrease the average degree $d_{\text{ave}}(G)$. Thus, we should have

$$d_{\text{ave}}(G - v) = \frac{2(e(G) - d(v))}{|V| - 1} \geq d_{\text{ave}}(G) = \frac{2e(G)}{|V|}$$

which is equivalent to $d(v) \leq d_{\text{ave}}(G)/2$. So, when we stick, each vertex in the resulting graph H has minimum degree at least $d_{\text{ave}}(G)/2$. $\square$

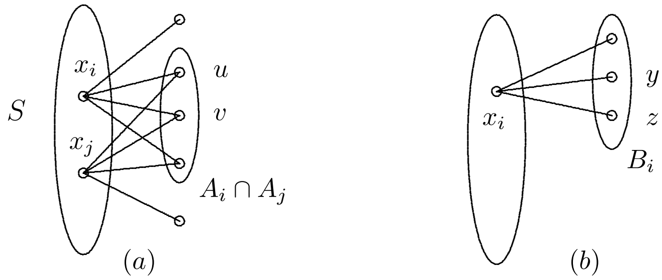


Fig. 2.2 (a) If u and v were non-adjacent, we would have an induced 4-cycle $\{x_i, x_j, u, v\}$. (b) If y and z were non-adjacent, then $(S \setminus \{x_i\}) \cup \{y, z\}$ would be a larger independent set.

Recall that a set of vertices in a graph is *independent* if no two of its vertices are adjacent. Let $\alpha(G)$ denote the largest number of vertices in such a set.

Proposition 2.8. *For every weakly C_4-free graph G on n vertices, we have*

$$\omega(G) \geq \frac{n}{\binom{\alpha(G)+1}{2}}.$$

Proof. Fix an independent set $S = \{x_1, \ldots, x_\alpha\}$ with $\alpha = \alpha(G)$. Let A_i be the set of neighbors of x_i in G, and B_i the set of vertices whose only neighbor in S is x_i. Consider the family $\mathcal{F}$ consisting of all α sets $\{x_i\} \cup B_i$ and $\binom{\alpha}{2}$ sets $A_i \cap A_j$. We claim that:

(i) each member of $\mathcal{F}$ forms a clique in G, and
(ii) the members of $\mathcal{F}$ cover all vertices of G.

The sets $A_i \cap A_j$ are cliques because G is weakly C_4-free: Any two vertices $u \neq v \in A_i \cap A_j$ must be joined by an edge, for otherwise $\{x_i, x_j, u, v\}$ would form a copy of C_4 as an induced subgraph. The sets $\{x_i\} \cup B_i$ are cliques because S is a maximal independent set: Otherwise we could replace x_i in S by any two vertices from B_i. By the same reason (S being a *maximal* independent set), the members of $\mathcal{F}$ must cover all vertices of G: If some vertex v were not covered, then $S \cup \{v\}$ would be a larger independent set.

Claims (i) and (ii), together with the averaging principle, imply that

$$\omega(G) \geq \frac{n}{|\mathcal{F}|} = \frac{n}{\alpha + \binom{\alpha}{2}} = \frac{n}{\binom{\alpha+1}{2}}. \qquad \square$$

Proposition 2.9. *Let G be a weakly C_4-free graph on n vertices, and $d = d_{\min}(G)$. Then, for every $t \leq \alpha(G)$,*

$$\omega(G) \geq \frac{d \cdot t - n}{\binom{t}{2}}.$$

Proof. Take an independent set $S = \{x_1, \ldots, x_t\}$ of size t and let A_i be the set of neighbors of x_i in G. Let m be the maximum of $|A_i \cap A_j|$ over all $1 \leq i < j \leq t$. We already know that each $A_i \cap A_j$ must form a clique; hence, $\omega(G) \geq m$. On the other hand, by the Bonferroni inequality (Exercise 1.37) we have that

$$n \geq \left|\bigcup_{i=1}^{t} A_i\right| \geq td - \sum_{i<j} |A_i \cap A_j| \geq td - \binom{t}{2} m \,,$$

from which the desired lower bound on $\omega(G)$ follows. □

Now we are able to prove Theorem 2.6.

Proof of Theorem 2.6. Let a be the average degree of G; hence, $a = 2|E|/n$. By Proposition 2.7, we know that G has an induced subgraph of average degree $\geq a$ and minimum degree $\geq a/2$. So, we may assume w.l.o.g. that the graph G itself has these two properties. We now consider the two possible cases.

If $\alpha(G) \geq 4n/a$, then we apply Proposition 2.9 with* $t = 4n/a$ and obtain

$$\omega(G) \geq \frac{(a/2) \cdot t - n}{\binom{t}{2}} = \frac{n}{\binom{4n/a}{2}} \,.$$

If $\alpha(G) \leq 4n/a$, then we apply Proposition 2.8 and obtain

$$\omega(G) \geq \frac{n}{\binom{\alpha(G)+1}{2}} \geq \frac{n}{\binom{4n/a+1}{2}} \,.$$

In both cases we obtain

$$\omega(G) \geq \frac{n}{\binom{4n/a+1}{2}} = \frac{a^2}{8n + 2a} \geq 0.1 \frac{a^2}{n} \,.$$ □

2.4 Zarankiewicz's problem

At most how many 1s can an $n \times n$ 0-1 matrix contain if it has no $a \times b$ submatrix whose entries are all 1s? Zarankiewicz (1951) raised the problem of the estimation of this number for $a = b = 3$ and $n = 4, 5, 6$ and the general problem became known as *Zarankiewicz's problem*.

It is worth reformulating this problem in terms of bipartite graphs. A bipartite graph with parts of size n is a triple $G = (V_1, V_2, E)$, where V_1 and V_2 are disjoint n-element sets of *vertices* (or *nodes*), and $E \subseteq V_1 \times V_2$ is the set of *edges*. We say that the graph contains an $a \times b$ *clique* if there exist an

* For simplicity, we ignore ceilings and floors.