Turán's Theorem and Coding Theory

Frank R. Kschischang
Dept. of Electrical and Computer Engineering
University of Toronto

May 16, 2012

1 Turán's Theorem

Let G be a simple graph with n vertices and e edges. If e is large, one would expect that G should contain many *cliques*, i.e., collections of mutually neighbouring vertices. A natural question arises: if G does not contain a $(k+1)$-clique (i.e., a clique of $k+1$ vertices), what is the largest possible value for e? Let us denote by $T(n,k)$ the largest possible number of edges in a $(k+1)$-clique-free simple graph with n vertices, and let us refer to any $(k+1)$-clique-free simple graph with n vertices having $T(n,k)$ edges as *extremal*. Clearly $T(n,1)=0$, and $T(n,k)$ must be a non-decreasing function of k.

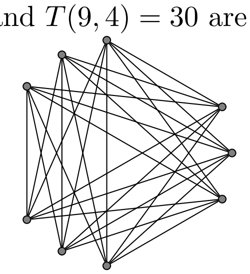
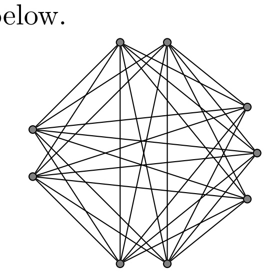
Turán's theorem, a fundamental result in extremal graph theory, provides an exact formula for $T(n,k)$, and a characterization of the extremal graphs.

Theorem 1 (Turán) *Let $n = qk + r$, where q and r are integers and $0 \le r < k$. Then*

$$T(n,k) = \frac{k-1}{2k}n^2 - \frac{r}{2}\left(1 - \frac{r}{k}\right),$$

achieved, uniquely, by the complete multipartite graph $K_{\underbrace{q,\ldots,q}_{k-r},\underbrace{q+1,\ldots,q+1}_{r}}$ having k vertex classes, r of them with $q+1$ vertices and the rest with q vertices.

A complete multipartite graph in which the number of elements in different vertex classes differs by at most one is known as a *Turán graph*, in connection with this theorem. For example, the graphs achieving $T(9,3) = 27$ and $T(9,4) = 30$ are shown below.

Before proving this theorem, let $G = (V, E)$. Let us write $\partial(v)$ for the degree of a vertex $v \in V$, i.e., for the number of edges of E incident on v. If E contains an edge incident on vertices u and v, let us write $uv \in E$, and call u and v *neighbours* in G. Let us write $u \asymp v$ if $uv \notin E$, i.e., if u and v are *not* neighbours in G.

Clearly $v \asymp v$ for all vertices v, and if $v \asymp w$ then $w \asymp v$ for all pairs of vertices v, w; thus the relation $\asymp$ is reflexive and symmetric. However, in a general graph G, it is *not* true that if $u \asymp v$ and $v \asymp w$ then $u \asymp w$, i.e., $\asymp$ is *not* transitive in general.

Now let $G = (V, E)$ be any simple graph. If we have a pair $u, v \in V$ with $u \asymp v$ and with $\partial(u) > \partial(v)$, then G can be modified to have more edges, without introducing a clique larger than any of the cliques in G. Simply delete vertex v (and all edges incident on v) and *clone* u, i.e., create a copy of u' of u, and include a new edge $u'w$ in E whenever uw is in E. Call the resulting graph $G' = (V', E')$, and note that $|V'| = |V|$. Since a clique cannot contain both u and u', any clique containing u' cannot be larger than a clique containing u. The number of edges in G' is given by

$$|E'| = |E| - \partial(v) + \partial(u) > |E|.$$

Thus in an extremal graph, non-neighbouring vertices must have equal degree.

A similar argument applies when a non-neighbour has the *same* degree as a pair of neighbouring vertices of that same degree. Suppose we have $G = (V, E)$ without a k-clique and a triple $u, v, w \in V$ with $u \asymp v$, $u \asymp w$, $vw \in E$ and $\partial(u) = \partial(v) = \partial(w)$. Again G can be modified to have more edges, without introducing any cliques larger than those present in G. Simply delete vertices v and w and clone u *twice*. By the same reasoning as in the previous paragraph, no large cliques are introduced by this procedure. In the resulting graph $G' = (V', E')$, we have $|V'| = |V|$ and

$$|E'| = |E| - (\partial(v) + \partial(w) - 1) + 2\partial(u) = |E| + 1.$$

The previous two paragraphs imply that, in an extremal graph (a) one cannot find a pair u, v with $u \asymp v$ and $\partial(u) \neq \partial(v)$ and (b) if $u \asymp v$ and $v \asymp w$, then $u \asymp w$, i.e., the relation $\asymp$ is transitive, and hence is an equivalence relation.

An extremal graph is thus multipartite and complete: the vertices can be partitioned into the equivalence classes of $\asymp$, and each vertex in a given class must be a neighbour of every vertex not in that class. (This automatically ensures that the degree of each vertex within a given class is the same.) Note that a complete multipartite graph with k vertex classes contains a k-clique (simply take k vertices from distinct classes), but no $(k+1)$-clique (since every set of $k+1$ vertices must, by the pigeonhole principle, contain at least two vertices from the same class).

Now, of the complete multi-partite graphs on n vertices not having a $(k+1)$-clique, which have the most edges? Note that an extremal $(k+1)$-clique-free graph must contain a k-clique, otherwise adding an edge would not create $(k+1)$-clique. Thus we can restrict our attention to complete multipartite graphs with exactly k vertex classes $V_1, \ldots, V_k$.

By definition $\sum_{i=1}^{k} |V_i| = n$. The degree of each vertex in V_i is given by $n - |V_i|$, and hence the total number of edges in the graph is given by

$$|E| = \frac{1}{2}\sum_{i=1}^{k} |V_i|\,(n - |V_i|) = \frac{1}{2}\left(n^2 - \sum_{i=1}^{k} |V_i|^2\right).$$

To maximize $|E|$, we must solve the following optimization problem: we must choose positive integers $|V_1|, \ldots, |V_k|$ so as to minimize $\sum_{i=1}^{k} |V_i|^2$, subject to $\sum_{i=1}^{k} |V_i| = n$. Without the integer constraint, a Lagrange multipliers approach would easily show that the optimal solution is to make all of the $|V_i|$'s equal. The actual solution makes them as equal as possible, while still satisfying the integer constraint.

Suppose for some i, j, we have $|V_i| \geq |V_j| + 2$. Modify G to G' by deleting a vertex from V_i and adding one to V_j; and let $|V_i'| = |V_i| - 1$, $|V_j'| = |V_j| + 1$, and $|V_k'| = |V_k|$ when $k \neq i, j$. Then

$$\begin{aligned}
\sum_{i=1}^{k} |V_i|^2 - \sum_{i=1}^{k} |V_i'|^2 &= |V_i|^2 + |V_j|^2 - (|V_i| - 1)^2 - (|V_j| + 1)^2 \\
&= 2(|V_i| - |V_j| - 1) \\
&> 0.
\end{aligned}$$

Thus G' would have more edges than G. It follows that, in an extremal configuration, the $|V_i|$'s must be nearly equal: any $|V_i|$ can differ from any $|V_j|$ by at most one.

The extremal graph for a given n and k is now completely determined: it is a complete k-partite graph with vertices partitioned into nearly equally sized classes. Let q and r be integers so that $n = kq + r$ and $0 \leq r < k$. Then $k - r$ classes contain q vertices and r classes contain $q + 1$ vertices. It is now easy to count the number of edges; we find

$$|E| = \frac{1}{2}\left(n^2 - (k - r)q^2 - r(q + 1)^2\right),$$

which simplifies (after substituting $q = (n - r)/k$) to the expression given in Theorem 1.

Theorem 1 is often used in a slightly weaker form by observing that $T(n, k) \leq (k - 1)n^2/(2k)$ for any choice of n and k. From this, the following Lemma immediately follows.

Lemma 1 *A simple graph with n vertices and e edges must contain a $(k + 1)$-clique if*

$$e > \left(1 - \frac{1}{k}\right)\frac{n^2}{2}.$$

This guarantee—that a clique of a certain size must exist under some conditions—is very useful for proving the existence of certain error-correcting codes, as we shall see next.

2 Codes are Cliques

As a warm-up, let d_H denote Hamming distance in the vector space $\mathbb{F}_q^n$. Consider the graph $G = (V, E)$ with q^N vertices in which $V = \mathbb{F}_q^N$. Allow $uv \in E$ if and only if $d_H(u, v) \geq d$, i.e., if the Hamming distance between the corresponding vectors is at least d. A *clique* in G is therefore a set of vectors whose pairwise Hamming distance is at least d, i.e., a code of length N over $\mathbb{F}_q$ of minimum Hamming distance at least d.

Note that G is regular: the degree of each vertex is

$$\partial(v) = \sum_{i=d}^{N} \binom{N}{i} (q-1)^i = q^N - \sum_{i=0}^{d-1} \binom{N}{i} (q-1)^i = q^N - V_{d-1},$$

where V_{d-1} denotes the volume of a Hamming ball of radius $d-1$ in $\mathbb{F}_q^N$. It follows that the number of edges $|E|$ is given by

$$|E| = \frac{1}{2} q^N \partial(v) = \frac{1}{2}(q^{2N} - q^N V_{d-1}).$$

According to Lemma 1, a clique of size $K+1$ in G (equivalently, a code with $K+1$ codewords of length N and minimum Hamming distance d) certainly exists if $|E| > \left(1 - \frac{1}{K}\right) \frac{q^{2N}}{2}$, i.e., if

$$\frac{1}{2}(q^{2N} - q^N V_{d-1}) > \frac{1}{2}\left(1 - \frac{1}{K}\right) q^{2N}$$

or

$$1 - \frac{V_{d-1}}{q^N} > 1 - \frac{1}{K}$$

or

$$K < \frac{q^N}{V_{d-1}},$$

which is a statement of the Gilbert-Varshamov bound.

Now consider a set X and a distance function $\rho : X \times X \to \mathbb{Z}^{\geq 0}$. Let $V_r(x)$ denote the volume of the ball of "radius" r centered at x, i.e.,

$$V_r(x) = |\{x' \in X : \rho(x, x') \leq r\}|.$$

As above, consider the graph $G = (V, E)$ with $V = X$, and $uv \in E$ if and only if $\rho(u, v) \geq d$. The degree of a vertex x is given by $|X| - V_{d-1}(x)$, and hence the total number of edges in the graph is given by

$$\begin{aligned} |E| &= \frac{1}{2} \sum_{x \in X} (|X| - V_{d-1}(x)) \\ &= \frac{|X|}{2} \left(|X| - \overline{V}_{d-1}\right), \end{aligned}$$

where

$$\overline{V}_{d-1} = \frac{1}{|X|} \sum_{x \in X} V_{d-1}(x)$$

denotes the *average* volume of a $(d-1)$-ball.

According to Lemma 1, a clique of size $K+1$ in G (equivalently, a code with $K+1$ codewords from X and minimum ρ-distance d) certainly exists if $|E| > (1-1/(K))|X|/2$, i.e., if

$$\frac{|X|}{2}\left(|X| - \overline{V}_{d-1}\right) > \frac{|X|}{2}\left(1 - \frac{1}{K}\right)|X|$$

or

$$1 - \frac{\overline{V}_{d-1}}{|X|} > 1 - \frac{1}{K}$$

or

$$K < \frac{|X|}{\overline{V}_{d-1}},$$

which is a statement of the so-called *generalized* Gilbert-Varshamov bound.

3 Notes

The content of this article is based on the work of Tolhuizen [1]. Turán's paper [2] was published in 1941 and is regarded as the starting-point of extremal graph theory. Many proofs of Turán's theorem are known; for example, the award-winning paper of Aigner [3] gives six proofs. A particularly short proof appears in [4, Ch. 4].

References

[1] L. M. G. M. Tolhuizen, "The generalized Gilbert-Varshamov bound is implied by Turán's Theorem," *IEEE Trans. Info. Theory*, vol. 43, pp. 1605–1606, Sept. 1997.

[2] P. Turán, "On an extremal problem in graph theory" (in Hungarian), *Math. Fiz. Lapok*, vol. 48, pp. 436–452, 1941.

[3] M. Aigner, "Turán's graph theorem," *Amer. Math. Monthly*, vol. 102, pp. 808–816, 1995. (Winner of a 1996 Lester R. Ford award for an article of expository excellence published in *The American Mathematical Monthly*.)

[4] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*, 2nd edition. Cambridge University Press, 2001.