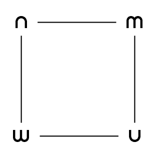
1 Texting with weird alphabets

Suppose you are texting your friend, but your screen is broken, so you can only use the letters mnuw. In order to make things simpler, you agree on a code for what messages different letters represent; for instance, m might mean "let's get breakfast", n might mean "let's get lunch", u might mean "let's get dinner", and w might mean "let's get midnight snack". Or alternately, if you want to communicate more than 4 distinct ideas, you could agree on a code that matches the 16 possible pairs of letters to 16 meanings. More generally, if you agree on a set of messages assigned to strings of letters of length n, then you can communicate up to 4^n distinct messages.

But now suppose you're texting me. Your phone is still broken, so you can still only use the letters mnuw. However, I'm really bad at phones; in particular, I often hold my phone upside down, and don't notice. This means that if you text me an m, I might read it as w, and vice versa; similarly, I confuse n and u.

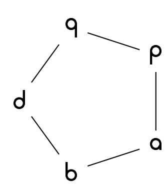
In this case, we'll be in trouble if we assign the same four messages to the four letters; if we do that, and you text me a w, I might try to meet you at midnight instead of at lunch. In fact, the pigeonhole principle shows that we can't assign more than two messages to these four letters: once we assign more than two, we will definitely be using two letters that I get confused, and we'll be in trouble. However, we can assign two messages, by simply restricting our attention to a subset of the four letters; there are several such options, such as n, m, or w, u. Similarly, if we assign messages to strings of length n, then we can send up to 2^n messages.

One way of representing the information of which letters I get confused is in terms of a so-called *distinguishability graph:* the vertices of the graph are the letters we're using, and there's an edge between two letters if I *can* tell them apart (so a non-edge represents confusability). In this example, the distinguishability graph would be



This feels a bit silly for such a simple example, but it will become important soon. Note that a set of letters we can restrict to is precisely a pair of letters in this graph that are connected by an edge. More generally, if we had a distinguishability graph with a subset of size r that looks like a complete graph on r vertices (this is called a *clique* in the graph), then we could restrict to those r letters and send r distinguishable messages unambiguously.

A more complicated example of a distinguishability graph is the following:



How many messages can we use with this set of five letters? Again, by pigeonhole, we can't do better than 2 (and in fact, by looking at the graph, we see that it has no clique subgraph of size 3 or more). Again, we get that with strings of length n, we can send up to 2^n messages.

However, something astonishing happens in this example: we can actually do *better* than 2^n. Specifically, using strings of length 2, we can actually send 5 (which is more than 4) messages: we restrict ourselves to the strings aa, bd, dp, qb, pq. Then it is a simple check that I will always be able to distinguish any pair of these strings; for some pairs, I'll confuse the first letter, but I'll be able to tell apart the second letter, while for other pairs the reverse will happen. Thus, these five strings do indeed allow us to encode five messages unambiguously, so that I will never confuse them. This also implies that if we send strings of length $2k$, then we will be able to encode 5^k messages. Thus, in a string of length n, we can encode roughly $(\sqrt{5})^n$ messages.

As both of these examples show, the number of messages we can send with strings of length n is roughly C^n, for some constant C. This makes sense, since we can always scale up multiplicatively: if we have r messages of length n, then by appending all pairs, we get r^2 messages of length $2n$. This motivates the following definition:

Definition (Shannon, 1956)**.** Given a graph G, its *Shannon capacity*, denoted $\Theta(G)$, is defined as the number of messages "per letter" that we can send using G. More precisely, it is the limit of $(r_n)^{1/n}$, where r_n is the number of messages that we can send using strings of length n.

Then what we saw above is that $\Theta(C_4) \geq 2, \Theta(C_5) \geq \sqrt{5}$, where C_n denotes a cycle graph of length n. We also saw above that $\Theta(G) \geq \omega(G)$, where $\omega(G)$ is the *clique number* of G—the size of the largest clique in G; this follows just because a clique of some size in G allows us to send that many messages using a single letter.

However, as in the case of C_5, we saw that sometimes we can do strictly better than the clique number, namely doing strictly better by sending longer messages. Maybe we can also do better in the case of C_4? As it turns out, we can't:

Theorem. $\Theta(C_4) = 2$.

Proof. We already proved that $\Theta(C_4) \geq 2$, so we just need to prove that $\Theta(C_4) \leq 2$. For that, we will cover C_4 by two *independent sets*, namely subsets of the vertices that contain no edges. In the example above, these independent sets will be red (R), consisting of m and ш, and blue (B), consisting of ∩ and ∪. Now, suppose that n is some integer, and that we have a set S of unambiguous messages of length n. We would like to show that $|S| \leq 2^n$. For each string in S, we form an auxiliary string as follows: for each letter in the string, we forget which letter it actually is, and only write down its color. In so doing, we get a string of length n in the letters R, B. Moreover, suppose that two distinct strings in S produce the same color-string. Then in each coordinate, they must have either the same letter or two confusable letters, since R and B were both independent sets. Since we assumed that all strings in S were not confusable, these two strings must agree in every coordinate, so they must actually be equal.

This thus gives us an injection from S to the set of all strings in the letters R, B. Since there are exactly 2^n such strings, we get that $|S| \leq 2^n$. □

In fact, the exact same proof shows that if we can cover G by k independent sets, then $\Theta(G) \leq k$. In the case of C_5, that shows us that $\Theta(C_5) \leq 3$. Moreover, as the above names of the independent sets suggest, the minimal number of independent sets it takes to cover a graph has a special name: it is the chromatic number $\chi(G)$, namely the minimal number of colors necessary to properly color G. So we have just proven the following theorem.

Theorem. $\omega(G) \leq \Theta(G) \leq \chi(G)$.

For many graphs, ω and χ are equal; for instance, this happens for all graphs on at most 4 vertices. This is what makes the 5-cycle so special: it is the smallest graph with $\omega(G) < \chi(G)$, which means that it is the smallest graph for which the above theorem does not determine Θ. Shannon himself proved this theorem when he first defined the Shannon capacity, and ended that paper with an open question, which was to determine $\Theta(C_5)$. Astonishingly, solving this problem took more than 20 years, and was eventually done by Lovász in 1978; by the end of this course, we will see that proof. But in the meantime, all we know is that $\sqrt{5} \leq \Theta(C_5) \leq 3$.

2 The OR product

In order to formalize the definition of the Shannon capacity, we need to first define a very important operation on graphs, which is called the OR product.

Definition. Given two graphs G, H, their OR product, denoted $G \cdot H$, is a graph with vertex set

$$V(G \cdot H) = V(G) \times V(H)$$

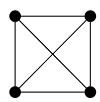
In other words, a vertex of $G \cdot H$ is an ordered pair, consisting of a vertex of G and a vertex of H. The edges of $G \cdot H$ are given by

$$E(G \cdot H) = \{((u,v),(u',v')) : (u,u') \in E(G) \text{ or } (v,v') \in E(H)\}$$

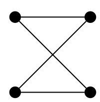
In other words, two ordered pairs are connected by an edge if and only if the restriction to at least one of the two coordinates is connected by an edge. Note that this is the mathematical usage of "or", which means that it is really "and/or"—both can be connected, but at least one must be.

Finally, given a positive integer n, we will denote by G^n the nth OR power of G, namely $\underbrace{G \cdot G \cdots G \cdot G}_{n \text{ times}}$.

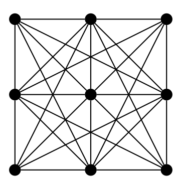
Example. If $G =$, then $G \cdot G$ is



On the other hand, if H is the two-vertex graph with no edge, , then $G \cdot H$ is



Finally, if $P_3 =$, then $P_3 \cdot P_3$ is



The reason we care about OR products and OR powers is the following proposition.

Proposition. *If G is the distinguishability graph of some set of letters, then a clique in G^n precisely corresponds to a collection of unconfusable strings of length n.*

Proof. First of all, a vertex of G^n is indeed precisely a string of length n consisting of the letters defining G. Moreover, two vertices in G^n are connected by an edge if and only if there is at least one coordinate in which they are connected (in G), which precisely means that there is at least one coordinate where they are distinguishable. Thus, a clique in G^n is precisely a collection of strings of length n, any two of which are distinguishable in at least one coordinate. This is precisely our previous notion of distinguishable strings, as desired. □

This proposition immediately tells us a very important fact, namely that our notion of "the most messages we can send with length n" is exactly $\omega(G^n)$. Thus, we can restate formally our definition from yesterday:

Definition (Shannon, 1956)**.** Given a graph G, its Shannon capacity, $\Theta(G)$, is

$$\Theta(G) = \lim_{n \to \infty} [\omega(G^n)]^{1/n}$$

An important property of OR products is the following lemma:

Lemma. *If G, H are graphs, then*

$$\omega(G \cdot H) \geq \omega(G) \cdot \omega(H)$$

Proof. Let C be a maximal clique in G, and D a maximal clique in H. Then consider $C \times D \subseteq V(G \cdot H)$. First of all,

$$|C \times D| = |C| \cdot |D| = \omega(G) \cdot \omega(H)$$

Moreover, $C \times D$ is a clique in $G \cdot H$. To see this, consider two distinct vertices $(u, v), (u', v') \in C \times D$. Since they're distinct, at least one of the pairs u, u' and v, v' contains distinct elements. Since these are vertices from a clique in G or H, they are connected by an edge. Thus, (u, v) and (u', v') are connected by an edge. So we have found a clique in $G \cdot H$ of size $\omega(G) \cdot \omega(H)$, which proves what we want. □

By applying this lemma n times, we find that

$$\omega(G^n) = \omega(G \cdot G \cdots G \cdot G) \geq \omega(G) \cdot \omega(G) \cdots \omega(G) \cdot \omega(G) = \omega(G)^n$$

and thus we conclude that

Corollary. $\Theta(G) \geq \omega(G)$.

We proved this yesterday intuitively, but it's good that our formal notion of Shannon capacity seems to work. In particular, the fact that $\omega(G^n) \geq \omega(G)^n$ precisely tells us what we saw yesterday, that if we have some number k of unambiguous letters, then they define k^n unambiguous strings of length n.

Yesterday, we also saw that $\Theta(G) \leq \chi(G)$; in order to prove it formally, we need the following very important lemma (that will also be very useful in the future).

Lemma. *Suppose that for every graph G, we have some real number $r(G)$, that satisfies the properties that for all graphs G, H*

1. $r(G) \geq \omega(G)$
2. $r(G \cdot H) \leq r(G) \cdot r(H)$

Then $\Theta(G) \leq r(G)$.

Proof. For any integer n, we have that

$$\omega(G^n) \leq r(G^n) = r(G \cdot G \cdots G \cdot G) \leq r(G) \cdot r(G) \cdots r(G) \cdot r(G) = r(G)^n$$

where the first inequality comes from property (1), and the second inequality comes from property (2). This implies that

$$[\omega(G^n)]^{1/n} \leq [r(G)^n]^{1/n} = r(G)$$

and thus

$$\Theta(G) = \lim_{n \to \infty} [\omega(G^n)]^{1/n} \leq \lim_{n \to \infty} r(G) = r(G)$$

□

Corollary. $\Theta(G) \leq \chi(G)$.

Proof. We check that χ satisfies properties (1) and (2). First, we already saw yesterday that $\omega(G) \leq \chi(G)$ for all G; this can also be seen by observing that it takes m colors to properly color a complete graph on m vertices, so it will take at least $\omega(G)$ colors to properly color G. For property (2), we need to check that

$$\chi(G \cdot H) \leq \chi(G) \cdot \chi(H)$$

for all graphs G, H. Indeed, suppose we have a coloring c_1 of G, namely a function $c_1 : V(G) \to C_1$ that assigns to each vertex some color in the set C_1. Similarly, suppose we have a coloring $c_2 : H \to C_2$. Finally, we may assume that these are minimal colorings, namely that $|C_1| = \chi(G), |C_2| = \chi(H)$. Then we can color $G \cdot H$ by declaring

$$c((u, v)) = (c_1(u), c_2(v)) \in C_1 \times C_2$$

This yields a function $c : V(G \cdot H) \to C_1 \times C_2$, and we claim that it's a proper coloring. Indeed, if we have an edge in $G \cdot H$ between (u, v) and (u', v'), then there must be an edge either between u, u' or v, v' (or both). Then since c_1, c_2 are proper colorings, we have that either $c_1(u) \neq c_1(u')$ or $c_2(v) \neq c_2(v')$ (or both). But in particular, $c((u, v)) \neq c((u', v'))$, so this is indeed a proper coloring. Moreover, the number of colors we used is

$$|C_1 \times C_2| = |C_1| \cdot |C_2| = \chi(G) \cdot \chi(H)$$

Thus, since $\chi(G \cdot H)$ is the minimal number of colors we need to properly color $G \cdot H$, and since we can color it with $\chi(G) \cdot \chi(H)$ colors, we find that

$$\chi(G \cdot H) \leq \chi(G) \cdot \chi(H)$$

This proves property (2), so the previous lemma guarantees that $\Theta(G) \leq \chi(G)$. $\square$

3 Orthonormal representations

Before we start today, we need to review some basic properties of dot products in $\mathbb{R}^n$. Recall that an element $\mathbf{u}$ of $\mathbb{R}^n$ is just an n-tuple $(u_1, \dots, u_n)$ of real numbers.

Definition. Given two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, their *dot product* is defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n$$

Note that the dot product of two vectors is a real number.

Definition. The *length* of a vector $\mathbf{u}$ is defined as

$$\|\mathbf{u}\| = \sqrt{u_1^2 + \cdots + u_n^2}$$

Note that $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$. If a vector has length 1, then it's called a *unit* (or *normal*) vector.

Proposition. *If two vectors* $\mathbf{u}, \mathbf{v}$ *have an angle* θ *between them, then*

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \cdot \|\mathbf{v}\| \cos\theta$$

In particular, two vectors form an angle of 90° *if and only if* $\mathbf{u} \cdot \mathbf{v} = 0$*. Such vectors are called orthogonal.*

Definition. Given a graph G, an *orthonormal representation* of G consists of some natural number n and, for every vertex $i \in V(G)$, a unit vector $\mathbf{u}_i \in \mathbb{R}^n$ such that if $(i, j) \in E(G)$, then $\mathbf{u}_i \cdot \mathbf{u}_j = 0$. In other words, we view the vertices of our graph as points in $\mathbb{R}^n$, and then we require that if two vertices are adjacent, then the corresponding unit vectors are orthogonal.

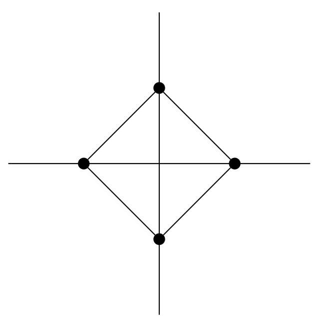
Note that we do *not* require anything in the case that i and j are not adjacent.

Example. A very simple example is that if G has n vertices, then we always have a simple orthonormal representation in $\mathbb{R}^n$: simply declare

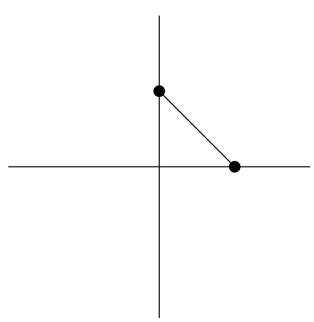
$$\mathbf{u}_i = (0, 0, \dots, 0, 1, 0, \dots, 0)$$

where the 1 is in position i. Since all of these vectors are orthogonal unit vectors, we certainly get an orthonormal representation of G, regardless of where its edges are.

Here is a slightly more interesting orthonormal representation of C_4 in $\mathbb{R}^2$:



The vertices are at the vectors $(\pm 1, 0), (0, \pm 1)$. Here is a different orthonormal representation of C_4:



In this case, we've put two vertices at $(1, 0)$ and two at $(0, 1)$, and the one edge drawn above actually represents all four edges.

One way of generalizing this last example is as follows. Suppose G is a graph, and $k = \chi(G)$. Then we can get an orthonormal representation of G in $\mathbb{R}^k$ as follows. First, color G with the colors $1, 2, \ldots, k$. Then, declare

$$\mathbf{u}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$$

where the 1 is in position ℓ if and only if vertex i is colored with color ℓ. This is indeed an orthonormal representation—first of all, all of these vectors are unit vectors, and second of all, if (i, j) is an edge of G, then i and j definitely received different colors, and thus $\mathbf{u}_i \cdot \mathbf{u}_j = 0$.

Intuitively, it seems that some of these orthonormal representations are more "wasteful" than others, in that the vectors we use are more "spread out". One way of capturing how spread out these are is the following quantity.

Definition (Lovász 1978)**.** Let $\mathbf{U} = \{\mathbf{u}_i\}_{i \in V(G)}$ be an orthonormal representation of a graph G in $\mathbb{R}^n$, and let $\mathbf{c} \in \mathbb{R}^n$ be some unit vector (which we call the *handle* of the representation). Then the Lovász ϑ-function of $\mathbf{U}$ and $\mathbf{c}$ is defined as

$$\vartheta(\mathbf{U}, \mathbf{c}) = \max_{i \in V(G)} \frac{1}{(\mathbf{c} \cdot \mathbf{u}_i)^2}$$

What is this actually measuring? Since $\mathbf{c}$ and $\mathbf{u}_i$ are unit vectors, we have that $\mathbf{c} \cdot \mathbf{u}_i = \cos\theta$, where θ is the angle between them. So in some sense, $\vartheta(\mathbf{U}, \mathbf{c})$ is measuring how wide a cone with axis at $\mathbf{c}$ has to be in order to contain all the vectors used in $\mathbf{U}$. However, I have no good intuition as to why we square or why we take reciprocals; it is just something that will make the proof work.

In order to figure out how "absolutely" spread out $\mathbf{U}$ is, we don't want to compare to a single handle $\mathbf{c}$, but to the best handle $\mathbf{c}$:

Definition. Given an orthonormal representation $\mathbf{U}$ of a graph G, we define

$$\vartheta(\mathbf{U}) = \min_{\substack{\mathbf{c} \in \mathbb{R}^n \\ \|\mathbf{c}\| = 1}} \vartheta(\mathbf{U}, \mathbf{c})$$

Finally, we want this to be a property of the graph, rather than a property of some orthonormal representation. So we also take the minimum over all orthonormal representations (in all dimensions):

$$\vartheta(G) = \min_{\mathbf{U} \text{ an orthonormal representation of } G} \vartheta(\mathbf{U})$$

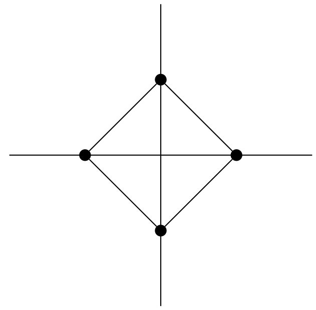
This is called the *Lovász ϑ-function* of G.

Example. Let's think about C_4 again. We can think of the orthonormal representation in $\mathbb{R}^4$, where our vertices are at $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)$. Then for any choice of a unit vector $\mathbf{c}$, we have that

$$\min_{i \in V(C_4)} |\mathbf{c} \cdot \mathbf{u}_i| \geq \frac{1}{2}$$

Intuitively, the reason for this is that the vector that will minimize $\mathbf{c} \cdot \mathbf{u}_i$ for all i is the one in which all coordinates are equal. The unit-length $\mathbf{c}$ with this property is $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, which indeed gives $\mathbf{c} \cdot \mathbf{u}_i = \frac{1}{2}$. So we can conclude that for this representation, we have $\vartheta = 1/(1/2)^2 = 4$.

On the other hand, if we look again at our first orthonormal representation,



then we see that for every unit vector $\mathbf{c} \in \mathbb{R}^2$ we pick, there will be at least one of these vertices that forms an angle of at most 45°. So we can conclude that for this representation, we have

$$\vartheta = \frac{1}{(\cos 45°)^2} = \frac{1}{(1/\sqrt{2})^2} = 2$$

The key property we used in the example above is the following fact from linear algebra, which we won't prove:

Proposition. *If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are n orthonormal vectors in $\mathbb{R}^n$, and $\mathbf{c}$ is any unit vector in $\mathbb{R}^n$, then there is some i with*

$$|\mathbf{c} \cdot \mathbf{v}_i| \geq \frac{1}{\sqrt{n}}$$

This immediately implies an interesting bound on the ϑ function of a graph:

Theorem. $\vartheta(G) \leq \chi(G)$.

Proof. Recall that if $k = \chi(G)$, we have an orthonormal representation of G in $\mathbb{R}^k$, gotten by sending all of the vertices in each color class to one of the basis vectors $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$. Then the previous proposition implies that for any unit vector $\mathbf{c}$, there is some i such that $|\mathbf{c} \cdot \mathbf{e}_i| \geq 1/\sqrt{k}$. Therefore, if we denote this orthonormal representation by $\mathbf{U}_\chi$, then we get that

$$\vartheta(\mathbf{U}_\chi) = \frac{1}{(1/\sqrt{k})^2} = k = \chi(G)$$

Since $\vartheta(G)$ is a minimum over all orthonormal representations, we see that $\vartheta(G) \leq \chi(G)$. □

The reason this is interesting is that we will eventually use the Lovász ϑ-function to upper-bound the Shannon capacity, and this proposition guarantees that it improves on our current best upper bound of $\chi(G)$. (Moreover, one can also prove that $\vartheta(G) \leq \chi_f(G)$, so it even beats the fractional chromatic number.)

In order to prove that ϑ upper-bounds Θ, we will use the key lemma that we proved yesterday. We will do this in several steps.

The goal for today is to prove that the Lovász ϑ-function upper-bounds the Shannon capacity. For this, we will use the key lemma we proved two days ago; namely, in order to show that $\Theta(G) \leq \vartheta(G)$, we need to show that $\omega(G) \leq \vartheta(G)$ and that $\vartheta(G \cdot H) \leq \vartheta(G)\vartheta(H)$.

Proposition. *For any graph G, $\omega(G) \leq \vartheta(G)$.*

Proof. To prove this, we will need a key fact from linear algebra, which is basically a generalization of the Pythagorean theorem to higher dimensions. It says that if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are a collection of n orthonormal vectors in $\mathbb{R}^n$, and if $\mathbf{c}$ is any vector in $\mathbb{R}^n$, then

$$\|\mathbf{c}\|^2 = \sum_{i=1}^{n} (\mathbf{c} \cdot \mathbf{v}_i)^2$$

We won't prove this, but you should convince yourself that in the case $n = 2$, it is precisely the Pythagorean theorem. This also implies that if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is some collection of orthonormal vectors in $\mathbb{R}^n$ (not necessarily n of them), then $\|\mathbf{c}\|^2 \geq \sum_{i=1}^{k} (\mathbf{c} \cdot \mathbf{v}_i)^2$.

Now, to prove that $\vartheta(G) \geq \omega(G)$, we need to prove that for any orthonormal representation $\mathbf{U}$ of G and any handle $\mathbf{c} \in \mathbb{R}^n$, we have that $\vartheta(\mathbf{U}, \mathbf{c}) \geq \omega(G)$. For this, let $K \subseteq V(G)$ be a maximal clique in G. Then $\{\mathbf{u}_i\}_{i \in K}$ is an orthonormal collection in $\mathbb{R}^n$, since all pairs are adjacent, so all the vectors are orthogonal. Thus, using the above, we have that

$$1 = \|\mathbf{c}\|^2 \geq \sum_{i \in K} (\mathbf{c} \cdot \mathbf{u}_i)^2 \geq |K| \min_{i \in V(G)} (\mathbf{c} \cdot \mathbf{u}_i)^2 = \omega(G) \min_{i \in V(G)} (\mathbf{c} \cdot \mathbf{u}_i)^2$$

When we divide out, we get that

$$\omega(G) \leq \frac{1}{\min_{i \in V(G)} (\mathbf{c} \cdot \mathbf{u}_i)^2} = \max_{i \in V(G)} \frac{1}{(\mathbf{c} \cdot \mathbf{u}_i)^2} = \vartheta(\mathbf{U}, \mathbf{c})$$

Since this holds for all $\mathbf{U}, \mathbf{c}$, we conclude that $\omega(G) \leq \vartheta(G)$, as desired. $\square$

Proposition. *For any two graphs G, H, we have $\vartheta(G \cdot H) \leq \vartheta(G)\vartheta(H)$.*

Proof. First, we need an important operation on vectors, called the *tensor product*. Given two vectors $\mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n$, we define $\mathbf{u} \otimes \mathbf{v} \in \mathbb{R}^{mn}$ by

$$\mathbf{u} \otimes \mathbf{v} = (u_1 v_1, u_1 v_2, \ldots, u_1 v_n, u_2 v_1, \ldots, u_2 v_n, \ldots, u_m v_1, \ldots, u_m v_n)$$

In other words, the mn coordinates of $\mathbf{u} \otimes \mathbf{v}$ consist of all the pairwise products of a coordinate of $\mathbf{u}$ and a coordinate of $\mathbf{v}$. One of the reasons we care about the tensor product is that it interacts very nicely with dot products. Namely, for any $\mathbf{u}, \mathbf{w} \in \mathbb{R}^m, \mathbf{v}, \mathbf{x} \in \mathbb{R}^n$, we have the formula

$$(\mathbf{u} \otimes \mathbf{v}) \cdot (\mathbf{w} \otimes \mathbf{x}) = (\mathbf{u} \cdot \mathbf{w})(\mathbf{v} \cdot \mathbf{x})$$

Note that on the left-hand side, we have the dot product of two vectors in $\mathbb{R}^{mn}$, whereas on the right-hand side, we have the product of two real numbers (each of which is a dot product of vectors). To prove this this formula, we just write down both sides: the left-hand side is

$$(\mathbf{u} \otimes \mathbf{v}) \cdot (\mathbf{w} \otimes \mathbf{x}) = (\ldots, u_i v_j, \ldots) \cdot (\ldots, w_i x_j, \ldots) = \sum_{i=1}^{m} \sum_{j=1}^{n} u_i v_j w_i x_j$$

The right-hand side is

$$(\mathbf{u} \cdot \mathbf{w})(\mathbf{v} \cdot \mathbf{x}) = \left(\sum_{i=1}^{m} u_i w_i \right) \left(\sum_{j=1}^{n} w_i x_j \right) = \sum_{i=1}^{m} \sum_{j=1}^{n} u_i v_j w_i x_j$$

which proves the formula.

Now recall what we want to prove, namely that $\vartheta(G \cdot H) \leq \vartheta(G)\vartheta(H)$. Since the ϑ-function is defined as a minimum over all orthonormal representations, in order to prove an upper bound for $\vartheta(G \cdot H)$, it suffices to present an orthonormal representation of $G \cdot H$ for which we can upper-bound ϑ. For this, let's pick optimal orthonormal representations and handles for G and H; namely, we pick an orthonormal representation $\mathbf{U}$ of G in dimension m and $\mathbf{V}$ of H in dimension n, along with unit vectors $\mathbf{c} \in \mathbb{R}^m, \mathbf{d} \in \mathbb{R}^n$, such that

$$\vartheta(G) = \vartheta(\mathbf{U}, \mathbf{c}) \qquad \text{and} \qquad \vartheta(H) = \vartheta(\mathbf{V}, \mathbf{d})$$

Then, we construct an orthonormal representation of $G \cdot H$ as follows. Recall that a vertex of $G \cdot H$ is a pair (i, j) with $i \in V(G), j \in V(H)$. To this vertex, we assign the vector $\mathbf{u}_i \otimes \mathbf{v}_j \in \mathbb{R}^{mn}$, and we claim that this forms an orthonormal representation $\mathbf{U} \otimes \mathbf{V}$ of $G \cdot H$ in $\mathbb{R}^{mn}$. To see this, let's first check that each of these vectors are unit vectors. Indeed,

$$\|\mathbf{u}_i \otimes \mathbf{v}_j\| = \sqrt{(\mathbf{u}_i \otimes \mathbf{v}_j) \cdot (\mathbf{u}_i \otimes \mathbf{v}_j)} = \sqrt{(\mathbf{u}_i \cdot \mathbf{u}_i)(\mathbf{v}_j \cdot \mathbf{v}_j)} = \sqrt{1 \cdot 1} = 1$$

where we use the fact that $\mathbf{U}$ and $\mathbf{V}$ are orthonormal representations. Next, we need to check the orthogonality condition: if (i, j) is adjacent to (i', j') in $G \cdot H$, then we need to make sure that $\mathbf{u}_i \otimes \mathbf{v}_j$ is orthogonal to $\mathbf{u}_{i'} \otimes \mathbf{v}_{j'}$. For this, we again use our formula:

$$(\mathbf{u}_i \otimes \mathbf{v}_j) \cdot (\mathbf{u}_{i'} \otimes \mathbf{v}_{j'}) = (\mathbf{u}_i \cdot \mathbf{u}_{i'})(\mathbf{v}_j \cdot \mathbf{v}_{j'})$$

Now, since (i, j) is adjacent to (i', j') in $G \cdot H$, then by the definition of the OR product, we must have that either i is adjacent to i' in G or that j is adjacent to j' in H (or both). But in either of those two cases, since we have that $\mathbf{U}, \mathbf{V}$ are orthonormal representations, we get that the right-hand side is 0. Thus, we indeed get that $\mathbf{u}_i \otimes \mathbf{v}_j$ is orthogonal to $\mathbf{u}_{i'} \otimes \mathbf{v}_{j'}$, and thus $\mathbf{U} \otimes \mathbf{V}$ is an orthonormal representation of $G \cdot H$ in $\mathbb{R}^{mn}$.

Next, we also pick a handle for this representation, and there is a natural guess, namely $\mathbf{c} \otimes \mathbf{d} \in \mathbb{R}^{mn}$. As above, $\mathbf{c} \otimes \mathbf{d}$ is a unit vector in $\mathbb{R}^{mn}$, so it can act as an handle. Finally, we want to evaluate $\vartheta(\mathbf{U} \otimes \mathbf{V}, \mathbf{c} \otimes \mathbf{d})$. For this, we use our formula one last time:

$$\begin{aligned}
\vartheta(\mathbf{U} \otimes \mathbf{V}, \mathbf{c} \otimes \mathbf{d}) &= \max_{(i,j) \in V(G \cdot H)} \frac{1}{((\mathbf{c} \otimes \mathbf{d}) \cdot (\mathbf{u}_i \otimes \mathbf{v}_j))^2} \\
&= \max_{(i,j) \in V(G \cdot H)} \frac{1}{(\mathbf{c} \cdot \mathbf{u}_i)^2 (\mathbf{d} \cdot \mathbf{v}_j)^2} \\
&= \left(\max_{i \in V(G)} \frac{1}{(\mathbf{c} \cdot \mathbf{u}_i)^2} \right) \left(\max_{j \in V(H)} \frac{1}{(\mathbf{d} \cdot \mathbf{v}_j)^2} \right) \\
&= \vartheta(\mathbf{U}, \mathbf{c}) \vartheta(\mathbf{V}, \mathbf{d}) \\
&= \vartheta(G)\vartheta(H)
\end{aligned}$$

where the last line uses the fact that $(\mathbf{U}, \mathbf{c})$ are optimal choices for G, and $(\mathbf{V}, \mathbf{d})$ are optimal choices for H. From this, we conclude that

$$\vartheta(\mathbf{U} \otimes \mathbf{V}) \leq \vartheta(\mathbf{U} \otimes \mathbf{V}, \mathbf{c} \otimes \mathbf{d}) = \vartheta(G)\vartheta(H)$$

since the left-hand side is the minimum over all choices of handle. Similarly, we now conclude that

$$\vartheta(G \cdot H) \leq \vartheta(\mathbf{U} \otimes \mathbf{V}) \leq \vartheta(G)\vartheta(H)$$

since $\vartheta(G \cdot H)$ is defined as the minimum over all orthonormal representations. This is precisely what we wanted to prove. □

These last two propositions, taken together with the key lemma from two days ago, proves the following crucial corollary:

Corollary (Lovász, 1978)**.** *For any graph G, $\Theta(G) \leq \vartheta(G)$.*

4 Back to C_5

Finally, we can state Lovász's Theorem:

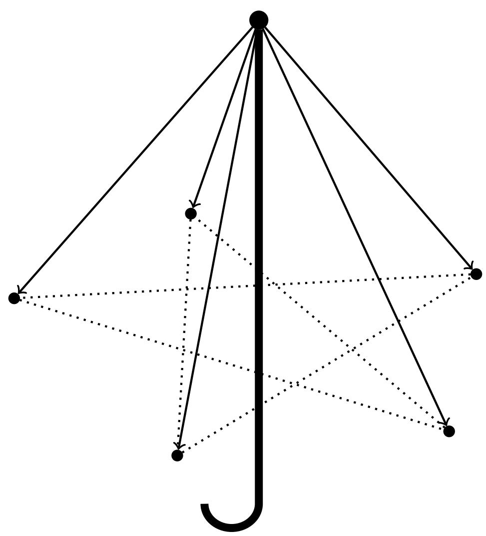
Theorem (Lovász, 1978)**.** $\Theta(C_5) = \sqrt{5}$.

Proof. We already proved on the first day of class that $\Theta(C_5) \geq \sqrt{5}$, so we just need to prove the reverse inequality. Since we've been building up the machinery of the Lovász ϑ-function, it's probably not surprising that we will prove that $\vartheta(C_5) \leq \sqrt{5}$. This will imply that

$$\sqrt{5} \leq \Theta(C_5) \leq \vartheta(C_5) \leq \sqrt{5}$$

and thus we have equalities throughout.

To prove that $\vartheta(C_5) \leq \sqrt{5}$, we will present a specific orthonormal representation of C_5 in $\mathbb{R}^3$, called the *Lovász umbrella* construction. For this, fix a handle $\mathbf{c}$, and imagine that your five vertices are on the boundary of an umbrella that we are slowly opening around the handle, with the tip of the umbrella at the origin. Keep track of the star (which is just C_5) that these five vertices define. If we let the umbrella open all the way, then the angle between two vertices connected by the star edges will be 144°. So at some point during the process, we must have all these angles be exactly 90°; this point at the process defines an orthonormal representation of C_5.



Now we need to do a bit of geometry to figure out exactly what the ϑ-function of this representation is. When the angles at the top of the umbrella are 90°, then every diagonal of the pentagon is the hypotenuse of an isosceles right triangle with side length 1. So every diagonal of the pentagon has length $\sqrt{2}$.

On yesterday's homework, you proved that if a regular pentagon has side length s, then its diagonal has side length ϕs, where $\phi = \frac{1}{2}(\sqrt{5}+1)$ is the golden ratio. So this pentagon has side length $s = \sqrt{2}/\phi$. You also proved that the distance between a vertex and the center of the pentagon is $s\sqrt{\phi}/\sqrt[4]{5}$. So we find that the distance between each of our vectors and the center of the pentagon is

$$r = \frac{\sqrt{2}}{\phi}\frac{\sqrt{\phi}}{\sqrt[4]{5}} = \frac{\sqrt{2}}{\sqrt{\phi}\sqrt[4]{5}}$$

Now, the quantity we're interested in is $\mathbf{c} \cdot \mathbf{u}_i$, and since these are all unit vectors, this is just $\cos\theta$, where θ is the angle between the handle and each of the five vectors. By the definition of the cosine, applied to the right triangle whose hypotenuse is the vector and one of whose legs is the handle, we get that $\cos\theta$ is just the length of the handle up to the point where it intersects the plane of the pentagon. Using the Pythagorean theorem on this same right triangle, this quantity is just $\sqrt{1-r^2}$. So we find that

$$(\mathbf{c} \cdot \mathbf{u}_i)^2 = (\cos\theta)^2 = 1 - r^2 = 1 - \frac{2}{\phi\sqrt{5}} = \frac{\phi\sqrt{5}-2}{\phi\sqrt{5}} = \frac{\phi}{\phi\sqrt{5}} = \frac{1}{\sqrt{5}}$$

Thus, we conclude that for the Lovász umbrella representation, we have

$$\vartheta(\mathbf{U}, \mathbf{c}) = \max_{i \in V(C_5)} \frac{1}{(\mathbf{c} \cdot \mathbf{u}_i)^2} = \frac{1}{1/\sqrt{5}} = \sqrt{5}$$

and thus $\vartheta(C_5) \leq \sqrt{5}$, as desired. □

5 What's next?

To recap, Shannon himself managed to determine the Shannon capacity of any graph which has $\omega = \chi$, and it took more than 20 years until Lovász determined the Shannon capacity of the smallest graph without this property. More or less, this is still the state of the art. In particular, we still don't know the Shannon capacity of any odd cycle larger than 5, all of which have $\omega = 2, \chi = 3$. The ϑ-function gives an upper bound that shows that $\Theta(C_n)$ approaches 2 as n gets large, and Bohman-Holzman proved a lower bound that $\Theta(C_n) > 2$ for all n, but these are the best we know. I believe that the Bohman-Holzman construction gives the correct value, but in order to prove this, we would need some new ideas to get an upper bound.

There's also something very interesting to say about computing these various things. In this class, we've proven the sequence of inequalities

$$\omega(G) \leq \Theta(G) \leq \vartheta(G) \leq \chi_f(G) \leq \chi(G)$$

It turns out that computing ω, χ, and χ_f is so-called "NP-hard"; this means that, as far as we know, computing these things takes an insanely long time. How long is insanely long? For a graph with 1000 vertices (which is quite small from a computer's perspective), if we put all of our supercomputers on the task of computing one of these parameters, they wouldn't finish before the death of the sun (in about 5 billion years).

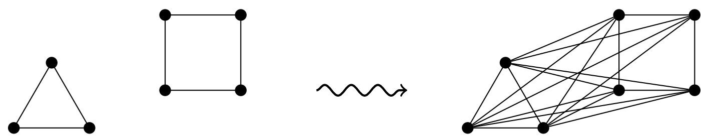
One of the most important breakthroughs that partially came out of Lovász's paper is the invention of *semidefinite programming,* which has become one of the most powerful tool in the theory of algorithms. In particular, it turns out that the ϑ-function (and any other so-called semidefinite problem) can be determined quickly (formally, in polynomial time). So despite being sandwiched between several NP-hard problems, ϑ turns out to not be hard to compute.

Finally, what about the Shannon capacity itself? The short answer is that we don't know. Some people believe that Θ might be *uncomputable*, meaning that determining it isn't just very hard—it might be actually impossible. I'm not sure I believe this conjecture, but if it is true, then it implies a crazy fact: there is some graph for which the axioms of math (called ZFC) cannot prove what its Shannon capacity is. So if this conjecture is true, more smart people like Lovász will not suffice for determining the Shannon capacity of every graphs: there are some for which this determination is literally impossible.

6 The Join of Graphs

Definition. Given two graphs G, H, their join is the graph $G + H$ gotten by placing G and H next to each other, and then connecting every vertex in G to every vertex in H.

Example. Here is the join $C_3 + C_4$:



From the perspective of communication, what does the join represent? Well, since there is an edge between every vertex of G and every vertex of H, the join precisely describes what happens when we put together two alphabets in such a way that I can always tell apart a letter of one from a letter of another. This interpretation immediately suggests a guess for $\Theta(G+H)$, namely $\Theta(G)+\Theta(H)$: a set of unconfusable messages in $G+H$ should just consist of taking together a set of unconfusable messages in G and in H.

Shannon himself formulated this guess as a conjecture, and he also managed to prove that $\Theta(G+H) \geq \Theta(G)+\Theta(H)$, and also proved equality for some special cases. This proof is not particularly hard, and the intuition is hopefully clear: to prove this, we legitimately can just construct a set of unconfusable messages in the vertices of $G+H$ of size $\Theta(G)+\Theta(H)$, basically by just putting together such sets for G and H.

However, in 1998, Noga Alon actually proved that this conjecture is *false.* Not only is it false, but it's somehow really really false:

Theorem (Alon, 1998)**.** *For every k, there exist graphs G, H with the property that $\Theta(G) \leq k, \Theta(H) \leq k$, but $\Theta(G+H) \geq k^2$.*

In fact, he proved an even stronger bound, which is a bit harder to state. We don't quite have time to prove this result, but we can see some of the big ideas in the proof.

First of all, one idea is to take $H = \overline{G}$, the complement graph of G. The reason is that this allows us to give a lower bound for $\Theta(G+H)$:

Proposition. *If G has n vertices, then $\Theta(G+\overline{G}) \geq \sqrt{n}$.*

Proof. It suffices to find a clique of size n in $(G+\overline{G})^2$. To do this, let's call the vertices of $G+\overline{G}$ $u_1, \ldots, u_n, v_1, \ldots, v_n$, where the u_i are the vertices of G, and the v_i are the vertices of $\overline{G}$. Then a clique in $(G+\overline{G})^2$ is given by $\{(u_1, v_1), \ldots, (u_n, v_n)\}$. Indeed, this is a clique for the following reason: for any i, j, we have that either u_i is adjacent to u_j in G, or not. If it is, then (u_i, v_i) is adjacent to (u_j, v_j) in $(G+\overline{G})^2$, since the first coordinate has an adjacency. If u_i is not adjacent to u_j, then v_i is adjacent to v_j in $\overline{G}$, and thus (u_i, v_i) is adjacent to (u_j, v_j) in $(G+\overline{G})^2$. Regardless, we get adjacency, so this is indeed a clique. Moreover, since G has n vertices, then this clique has size n, so

$$\Theta(G+\overline{G}) \geq \sqrt{\omega((G+\overline{G})^2)} \geq \sqrt{n}$$

□

The next part of Alon's proof is to pick a clever graph G, for which we can upper bound $\Theta(G)$ and $\Theta(\overline{G})$. The choice of G is not so important (though it also requires some ideas), but the main thing is Alon's new technique for constructing upper bounds for Θ. For this, he generalized Lovász's notion of an orthonormal representation to work for any *field*, which is basically a mathematical structure in which we can add, subtract, multiply, and divide. Crucially, in order to upper-bound both $\Theta(G)$ and $\Theta(\overline{G})$, Alon needed to do this sort of analysis over two different fields, and one can prove that this is necessary: a single upper bound like the ϑ-function cannot, by itself, give the necessary bounds for $\Theta(G)$ and $\Theta(\overline{G})$.