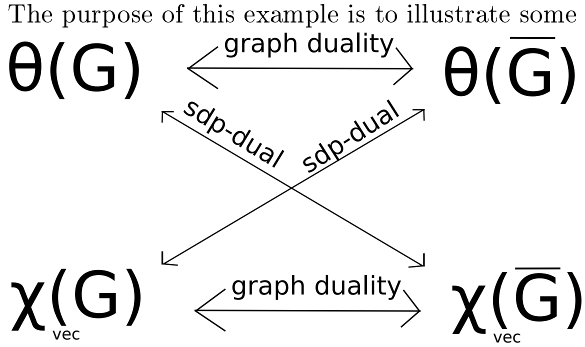
NOTES ON LOVASZ THETA

VICTOR BANKSTON

The purpose of this example is to illustrate some of the structure to $\vartheta(G)$.



Though the quantities above are scalars, they arise from arrangements of vectors. I will write down the vectors associated with each quantity above when G is the Petersen Graph.

Defining ϑ

First, we need a series of definitions to define ϑ:

Definition 1. Given a graph G, an orthonormal representation of G is a mapping $r : V(G) \to \mathbb{S}^n \subset \mathbb{R}^{n+1}$ (for some $n \in \mathbb{N}$) such that if $i \neq j \in V(G)$, with $i \not\sim j$, then $r(i) \perp r(j)$. (Be careful: a vertex is not adjacent to itself).

Note that each graph has at least one representation, where v maps the verticies each to its own orthonormal vector.

Definition 2. A valuation of an orthonormal representation $val(r)$ is

$$\min_{\psi} \max_{v \in V(G)} \frac{1}{\left(\psi^T r(v)\right)^2}$$

where ψ ranges over all unit vectors (of the target space of r).

Given an orthonormal representation, its valuation is how tightly it can be embedded into a cone around some vector (ψ).

Definition 3. Define $\vartheta(G)$ to be the minimum valuation over all orthonormal representations of G.

We can show that this minimum is actually attained. We will use Bolzano-Weirstrass. To see this, fix n, and consider the orthonormal representations of the form: $v : V(G) \to \mathbb{S}^{n-1} \subset \mathbb{R}^n$. Observe that $\vartheta(G)$ remains unchanged if we require that $\psi = (1, 0, 0, 0 \dots)$: These valuations are defined by an inner product, which will not change if we apply a fixed unitary U to every vector. Choose U to send $\psi \mapsto (1, 0, 0, 0 \dots)$.

Fixing ψ, take a sequence of orthonormal representations whose values converge to $\vartheta(G)$. Observe that these orthonormal representations themselves can be considered as bounded vectors of dimension $n \cdot V(G)$, by concatenating all $V(G)$ vectors of dimension n. By the Bolzano-Weirstrass theorem, these have a convergent subsequence, so there is an accumulation point, r_∞, which we must show is an orthonormal representation.

Our convergent subsequence of orthonormal representations gives rise to $V(G)$ convergent sequences of vectors. We must show that each sequence of vectors goes to a unit vector, and that when $i \not\sim j$, with $i \neq j$, we have $r_\infty(i)^T r_\infty(j) = 0$. Both of these are consequences of the fact that dot products are continuous: $0 = \lim_{n\to\infty} r_n(i)^T r_n(j) = (\lim_{n\to\infty} r_n(i))^T (\lim_{n\to\infty} r_n(j))$.

There is no claim that such optimal representations are unique.

Graphs

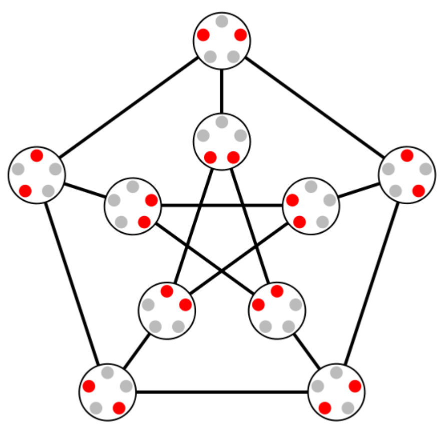
Definition 4. Define the Kneser Graph $k(n,r)$ to have $\binom{n}{r}$ verticies labeled by r-element subsets from a universe of size n. Two verticies are adjacent if their corresponding sets are disjoint. We assume that $n \geq 2r$

Kneser graphs are vertex and edge transitive. Given any two verticies (edges), there is an automorphism which sends one to the other.

Theorem 5. *If G is vertex and edge transitive, then $\vartheta(G)\vartheta(\overline{G}) = n$, and $\vartheta(G) = \frac{-n\lambda_n}{\lambda_1 - \lambda_n}$*

This powerful theorem was originally used to find $\vartheta(k(n,r))$. The proof of the theorem builds on the relations in the diagram.

Definition 6. The Petersen Graph, P, is the Kneser graph, $k(5,2)$.



We start with some graph properties.

Claim 7. The clique number of the kneser graph $\omega(k(n,r)) = \lfloor \frac{n}{r} \rfloor$, so $\omega(P) = 2$

A clique corresponds to a collection of disjoint sets.

Claim 8. The coloring number $\chi(k(n,r)) = n - 2r + 2$, so $\chi(P) = 3$

This was a big open problem for many years. The optimal coloring is the following: Order the elements of the universe $u_1, \ldots, u_n$, and divide them into 3 pieces with sizes $n-2r$, r and r. Let x be an r-set. If it intersects the first piece, color x with the color i, where $i=\min\{i \mid u_i \in x\}$. Otherwise, x is contained entirely in the last two pieces. These remaining verticies form a subgraph, where each vertex has a unique neighbor, and these can be colored with two colors.

Claim 9. The independence number of the kneser graph is $\alpha(k(n,r))=\binom{n-1}{r-1}$, so $\alpha(P)=4$

The collection of r-subsets which each contain u_1 is a set of this size. It isn't hard to show this is optimal.

Claim 10. The clique covering number is $q(k(n,r))=\left\lceil\frac{\binom{n}{k}}{\left\lfloor\frac{n}{k}\right\rfloor}\right\rceil$, $q(P)=5$

(!) (Claim found on Wolfram Mathworld.)

Claim 11. $\vartheta(k(n,r))=\binom{n-1}{r-1}$, and $\vartheta(k(n,r))=\frac{n}{r}$, so $\vartheta(P)=4, \vartheta\left(\overline{P}\right)=\frac{5}{2}$

This is proven by Theorem 5 and some tricky algebra, but this avoids (or at least obscures) creating explicit orthonormal representations, which is the point of this example.

Relations between Graph Constants

Theorem 12. $\alpha(G)\chi(G)\geq\mid V(G)\mid$

Each color is an independent set, and a proper coloring colors every vertex.

Theorem 13. *For any graph G, $\alpha(G)\leq\vartheta(G)\leq\chi\left(\overline{G}\right)=q(G)$*

In an orthonormal representation, an independent set, α, of G must be sent to a collection of pairwise independent vectors. For such vectors, it is easy to see that $\max_{v_i\in\alpha}\frac{1}{(\psi^T r(v_i))^2}$ is minimized when $\psi=\frac{\sum_{v_i\in\alpha} r(v_i)}{\sqrt{|V(G)|}}$ (when ψ is between all the vectors.) In this case, $\frac{1}{(\psi^T r(v_i))^2}=\mid\alpha\mid$, and this lower bound holds for all orthonormal representations. This shows $\alpha(G)\leq\vartheta(G)$.

Suppose we have clique cover of size $q(G)$. Define an orthonormal representation by choosing $q(G)$ pairwise orthonormal vectors. Send each clique to one of these vectors. This provides an explicit orthonormal representation with valuation $q(G)$. The minimum over all orthonormal representations may be less.

Orthonormal Representations

We start with the graph $\overline{k(n,r)}$ and construct an optimal orthonormal representation in dimension n, with orthonormal basis $u_1,\ldots,u_n$ (overloading the names of the basis elements with the elements of the universe) The choice is obvious: disjoint sets need to go to orthonormal vectors. Set $u_i^T r(v_j)=\frac{1}{\sqrt{r}}$ if $u_i\in v_j$, and 0 otherwise. Set $\psi=\frac{1}{\sqrt{n}}(1,1,\ldots,1)$. It is immediate that this is an orthonormal representation with valuation $\frac{1}{(\psi^T r(v_i))^2}=\frac{r\cdot n}{r^2}=\frac{n}{r}$. This O.R. spans a space of dimension 5.

Definition 14. Given an orthonormal representation, we can define the cost of a vertex to be $c(v)=\left(\psi_1^T(r_1(v_i))\right)^2$. This corresponds to the quantum-mechanical probability of measuring $r_1(v_1)$ when measuring from state ψ.

Theorem 15. *(Certification of Orthonormal Representations): if we have two orthonormal representations r_1, r_2 of G and $\overline{G}$ and for all $i \in V(G)$ we have $c_1(v_i) = \frac{1}{\vartheta}$, and we also have $\sum_i c_1(v_i) c_2(v_i) = 1$, then $\vartheta = \vartheta(G)$*

Proof. We have the explicit orthonormal representation r_1, so $\vartheta(G) \leq \vartheta$. For the other direction, we use an alternate definition ϑ, $\vartheta(G) = \max_{Rep(\overline{G})} \sum_i c(v_i)$.

$$\min_{r \in O.R.(G)} \max_i \frac{1}{c_r(v_i)} \leq \vartheta = \sum_i \vartheta c_1(v_i) c_2(v_i) = \sum_i c_2 v_i \leq \max_{r \in O.R.(\overline{G})} \sum_i c_r(v_i) = \vartheta(G)$$

□

The argument above also shows that certificates always exist.

The next definition is crucial, and describes the relationship between $\vartheta(G)$ and $\vartheta(\overline{G})$. From the physical perspective, this will relate bell inequalities of completely different experiments. Can this relation be found using the Sheaf Theory?

Definition 16. Given a non-empty closed convex set $P \subset \mathbb{R}_+^n$ with the property that $x \in P$ and $0 \leq x' \leq x$ then $x' \in P$, the antiblocker of P is

$$AB(P) = \left\{ x \in \mathbb{R}_+^n : y^T x \leq 1 \text{ for all } y \in P \right\}$$

The condition that $0 \leq x' \leq x \implies x' \in P$ implies that $AB(AB(P)) = P$.

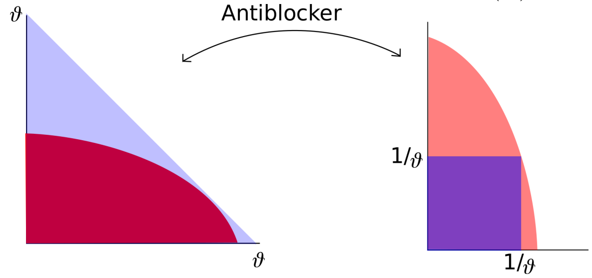
Example 17. Let $P_\vartheta = \left\{ y \in \mathbb{R}_+^n \mid \sum_{i=1}^n y_i \leq \vartheta \right\}$. Then $AB(P) = C_{\frac{1}{\vartheta}} = \left\{ x \in \mathbb{R}_+^n \mid \forall i, x_i \leq \frac{1}{\vartheta} \right\}$.

Proof. Let $y \in P_\vartheta, x \in C_{\frac{1}{\vartheta}}$. Then $y^T x = \sum_{i=1}^n y_i x_i \leq \frac{1}{\vartheta} \sum_{i=1}^n y_i \leq 1$. This shows $C_{\frac{1}{\vartheta}} \subset AB(P_\vartheta)$. Conversely, if $x \notin C_{\frac{1}{\vartheta}}$ for some $i \in V(G)$ $x_i > \frac{1}{\vartheta}$. Choose y such that $y_j = 0$ when $i \neq j$, and $y_i = \vartheta$. Then $y \in P_\vartheta$, and $x^T y > 1$, so $x \notin AB(P_\vartheta)$ □

Definition 18. $TH(G) = \left\{ (c(v_i), v_i \in V(G)) \in \mathbb{R}_+^{V(G)} \right\}$. These are assignable probabilities, which (claim) satisfy the hypotheses of definition 6. (Note, these probablities do not need to sum to 1. We allow that some experiments have outcomes which are disregarded. The problem is intractable otherwise.)

Theorem 19. $AB(TH(G)) = TH(\overline{G})$

These concepts provide a geometric description of two definitions of ϑ. ϑ is the maximal of a linear functional over $TH(G)$: $\vartheta(G) = \max_{\overline{O.R.}} \sum_{i=1}^{|V(G)|} c(v_i)$. This linear functional has hyperplanes as its level sets, and the optimal value corresponds to a level set which lies tangent to $TH(G)$. Thus, $\vartheta(G)$ is the smallest simplex S_ϑ such that $TH(G) \subset S_\vartheta$. If we take the antiblocker of this picture, we seek the reciprocal of the largest cube $C_{\frac{1}{\vartheta}}$ such that $C_{\frac{1}{\vartheta}} \subset AB(TH(G)) = TH(\overline{G})$. This explains the formula $\vartheta(G) = \min_{O.R.} \max_i \frac{1}{c(v_i)} = \frac{1}{\max_{O.R.} \min_i c(v_i)}$.



Next, we give an orthonormal representation of $\vartheta(P)$, which will certify the optimality of the orthonormal representation given at the begining of this section.

Assume a basis of size 10, $\{e_{s_1}, e_{s_2}, \ldots, e_{s_{10}}\}$ labeled by the $\binom{5}{2}$ subsets of the graph. Let $\psi = \frac{1}{\sqrt{10}}(1,1,\ldots,1)$. Finally, assume that we will have $e_{s_i}^T r(v_j) = x_{|s_i \cap v_j|}$. This is a plausible assumption, because it will result in vectors whose orthongonality relations are invariant with respect to the automorphism group of P.

The fact that intersecting sets must be sent to orthonormal vectors translates into the contstraint

$$x_0^2 + 3x_1^2 + 4x_0x_1 + 2x_1x_2 = 0$$

At the same time, we would like to minimize $\frac{10\cdot\left(x_2^2+6x_1^2+3x_0^2\right)}{(x_2+6x_1+3x_0)^2}$. According to Wolfram Alpha the minimum is 4, when $(x_0,x_1,x_2) = \left(1,-4-\sqrt{15},6+\sqrt{15}\right)$, or when $(x_0,x_1,x_2) = (a,b,c) = \left(1,\sqrt{15}-4,6-\sqrt{15}\right)$.

$$\begin{pmatrix} c & b & b & b & b & b & b & a & a & a \\ b & c & b & b & b & a & a & b & b & a \\ b & b & c & b & a & b & a & b & a & b \\ b & b & b & c & a & a & b & a & b & b \\ b & b & a & a & c & b & b & b & b & a \\ b & a & b & a & b & c & b & b & a & b \\ b & a & a & b & b & b & c & a & b & b \\ a & b & b & a & b & b & a & c & b & b \\ a & b & a & b & b & a & b & b & c & b \\ a & a & b & b & a & b & b & b & b & c \end{pmatrix}$$

These vectors span a space of dimension 6.

Remark 20. For any given graph G, it is not true that all optimal orthogonal representations have the same dimension. For example, there are two optimal orthonormal representations of C_4: $\{(0,1),(0,1),(1,0),(1,0)\}$ with handle $\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)$, and $\{(a,a,0),(a,-a,0),(a,0,a),(a,0,-a)\}$ with handle $(1,0,0)$.

Finally, we apply the certification theorem. For the O.R. above, each cost is $\frac{1}{4}$, and $\sum_{i=1}^{10}\frac{1}{4}\cdot\frac{2}{5} = 1$. Hence, the O.R. above is optimal. Similarly, our $O.R.$ of $\overline{K(n,r)}$ can be seen to be optimal.

Vector Colorings of Graphs

Definition 21. Given a graph G, we assign a unit vector to each vertex. This time, we would like adjacent verticies to be sent to vectors whose dot product is as negative as possible. If $\chi(G) = k$, then we can associate each color with a vector in the regular k-simplex in $\mathbb{R}^{k+1}$. Such vectors have inner product $\frac{-1}{k-1}$. In light of this, we define $\chi_{vec}(G) = \min\left\{k \mid v_i^T v_j = \frac{-1}{k-1}\right\}$

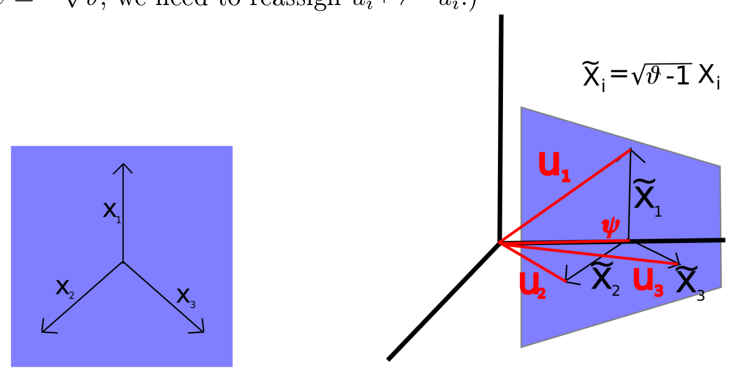
Theorem 22. $\chi_{vec}(G) = \vartheta\left(\overline{G}\right)$

This is proven by the fact that the two problems can be expressed as semidefinite-programming duals of one another. (The duality is between χ_{vec} and $\max_{\overline{O.R.}}\sum\left(\psi^T v_i\right)^2$.) Alternatively, there is a concrete way to move between optimal representations of the coloring problem and optimal representations for ϑ.

Proposition 23. *A vector coloring x_i is optimal (having value ϑ) iff $u_i = \frac{1}{\sqrt{\vartheta}}\left(\psi + \sqrt{\vartheta - 1}x_i\right)$ is an optimal Orthonormal representatio (also having value ϑ).*

Proof. If x_i is an optimal vector-coloring for G with coloring number ϑ, and ψ is some unit vector orthogonal to each x_i, then we obtain an orthonormal representation $u_i = \frac{1}{\sqrt{\vartheta}}\left(\psi + \sqrt{\vartheta - 1}x_i\right)$. First, observe that these are all unit vectors. Secondly, let $u_i \neq u_j$ correspond to non-adjacent verticies in $\overline{G}$, so that $x_i^T x_j = \frac{-1}{\vartheta - 1}$. Then $u_i^T u_j = \frac{1}{\vartheta}\left(1 + (\vartheta - 1)\frac{-1}{\vartheta - 1}\right) = 0$, so u is an orthogonal representation of $\overline{G}$. Also, we have $\frac{1}{(\psi^T u_i)^2} = \vartheta$.

Conversely, if we start with the orthonormal representation with value ϑ (so $\vartheta = \frac{1}{\left(u_i^T \psi\right)^2}$ for all i. We have not yet shown that it's always possible to achieve equality, but it can be seen from the antiblocker picture.) we can recover the coloring by $x_i = \frac{\sqrt{\vartheta}u_i - \psi}{\sqrt{\vartheta - 1}}$. Now, if $x_i \sim x_j$ in G, then $x_i^T x_j = \frac{-\sqrt{\vartheta}u_i^T\psi - \sqrt{\vartheta}u_j^T\psi + 1}{\vartheta - 1} = \frac{-1}{\vartheta - 1}$ so the coloring has value ϑ. Also, $x_i^2 = \frac{(\vartheta+1) - 2\sqrt{\vartheta}u_i \cdot \psi}{\vartheta - 1} = 1$. (There is a slight issue: if $u_i \cdot \psi = -\sqrt{\vartheta}$, we need to reassign $u_i \mapsto -u_i$.)



vector coloring of C_3 representation of $\overline{C}_3$(not normalized) □

Next, we will provide an optimal vector coloring of P. Assume a basis of size 5, and that we will map $\star\star\circ\circ\circ \mapsto (a, a, b, b, b)$, and extend this map by permutations of S_5. If x, y are two vector representations of intersecting sets, we would like to minimize

$$min_{a,b}\frac{x^T y}{\|x\|\,\|y\|} = \frac{4ab + b^2}{2a^2 + 3b^2}$$

The minumum occurs at $a = -3, b = 2$, and gives

$$\frac{x^T y}{\|x\|\,\|y\|} = \frac{-24 + 4}{18 + 12} = -\frac{2}{3} = \frac{-1}{\frac{5}{2} - 1}$$

Using numpy.linalg, we can find that these vectors span a space of dimension 4.

Finally, an optimal vector coloring of $\overline{P}$ can be found by assuming a basis of size 10 (the same basis we used for $\vartheta(P)$) and three variables, x_0, x_1, x_2. This gives us the optimization problem:

$$min_{x_0,x_1,x_2}\frac{x^T y}{\|x\|\,\|y\|} = \frac{x_0^2 + 3x_1^2 + 4x_0x_1 + 2x_1x_2}{3x_0^2 + 6x_1^2 + x_2^2}$$

The minimum (according to Wolfram) is found at $\left(-\frac{1}{\sqrt{18}}, \frac{1}{\sqrt{18}}, -\frac{1}{\sqrt{2}}\right)$ and gives

$$\frac{x^T y}{\|x\|\,\|y\|} = \frac{\frac{1}{18} + \frac{3}{18} - 4\frac{1}{18} - 2\frac{1}{6}}{3\frac{1}{18} + 6\frac{1}{18} + \frac{1}{2}} = \frac{-6}{18} = \frac{-1}{4-1}$$

This spans a space of dimension 5.

Problem 24. Since we have established that vector colorings correspond to orthonormal representations, we actually have 2 orthonormal representations of the Petersen Graph and 2 for its complement. Are these the same?

Quantum Measurements

We fix a finite-dimensional vector space, known as the state space. Pure quantum states will be given by unit vectors in this space.

> A projective measurement is described by an observable, M, a Hermitian operator on the state space of the system being observed. The observable has a spectral decomposition
>
> $$M = \sum_m m P_m$$
>
> where P_m is the projector onto the eigenspace of M with eigenvalue m. The possible outcomes of the measurement correspond to the eigenvalues, m, of the observable. Upon measuring the state $|\psi\rangle$, the probability of getting result m is given by
>
> $$\langle\psi|\, P_m \,|\psi\rangle$$
>
> Given that the outcome m occured, the state of the quantum system immediately after the measurement is $\frac{P_m|\psi\rangle}{\sqrt{p(m)}}$

(From Nielsen and Chuang)

(!) Naimark's Dilation Theorem states that all quantum measurements can be viewed as projective measurements on a larger system. The Spectral Theorem states that a Hermitian operator M always has a decomposition of the form

$$M = \sum_m m P_m$$

where m ranges over real numbers, and the ranges of the P_m's are pairwise orthogonal. Conversely, if we start with a collection of pairwise orthogonal vectors $\{|m\rangle\}_{m=1}^{n}$which span the state space, we can create a Hermitian operator which has those vectors as its eigenvectors: $M = \sum_{m=1}^{n} m\,|m\rangle\,\langle m|$. Thus, when specifying a PVM, we only need to supply an Orthonormal Basis.

Example 25. Polarizing sunglasses are an example of a PVM.

Light consists of an electric wave inducing a magnetic wave and vice versa. The direction of the electric wave determines the polarization of the light. A photon can be polarized in any 2-dimensional direction. Polarized lenses in sunglasses will let a photon through if it is polarized vertically, and will block it if it is polarized horizontally. The photon hitting the lens is a measurement. The state space is 2-dimensional, corresponding to the polarization directions. We are lucky in this case that polarization between horizontal and vertical directions has a direct physical meaning- that the light is polarized in a diagonal direction. Usually, this is not the case. Let $|v\rangle$ be the state of light which is polarized in the vertical direction and $|h\rangle$be the state of light polarized in the horizontal direction. When a vertically polarized photon hits a vertically polarized lens, it will surely pass through. We represent the observable for this lens by $L = 1 \cdot |v\rangle\langle v| + 0 \cdot |h\rangle\langle h|$. Suppose we have light that is polarized diagonally,

$$|\psi\rangle = \frac{1}{\sqrt{2}}|v\rangle + \frac{1}{\sqrt{2}}|h\rangle$$

The probability of the light passing through the lense is $(\langle v| + \langle h|)|v\rangle\langle v|(|v\rangle + |h\rangle) = \frac{1}{2}$, and if it does so, its state is $\frac{|v\rangle\langle v|\frac{1}{\sqrt{2}}(|v\rangle+|h\rangle)}{\sqrt{\frac{1}{2}}} = |v\rangle$. Observe that the probability that the light passes through the lens is the square of the inner product and the resultant state we're interested in. This is true in general.

Forward Example: The CHSH Inequality

There are two conceivable ways to use the machinery of ϑ to investigate contextuality. Usually we start with a collection of measurements and a linear functional on the outcomes of these measurements. By drawing the exclusion graph of the measurements, we can identify bounds (Bell Inequalities) on these linear functionals. We will see that these bounds are given by ϑ and α. Alternatively, we can go backwards and search for graphs which have a large gap between ϑ and α, then find orthonormal representations to realize these graphs as collections of quantum experiments. Ultimately, this is the direction that I would like to pursue, but we should start with the standard technique first.

As an example of the forward method, we will investigate the CHSH inequality. Imagine Alice and Bob, separated in time and space and unable to communicate. However, they share qubits which make up a quantum state. That is, these particles may be entangled, so the total state space is 4-dimensional. Alice may choose between two measurements A and A', and Bob may also choose between two measurements B and B', so there are 4 total measurements which may occur: $\{AB, A'B, AB', A'B'\}$. Each of Alice and Bob's local measurements yield an outcome of 0 or 1, so each total measurement has 4 possible outcomes.

(One formulation of) the CHSH inequality begins with such a scenario and, if Alice and Bob both choose their experiments uniformly at random, and assigns a value of 1 whenever the their outcomes agree and A, B is not chosen, and a value of -1 whenever their outcomes disagree and A, B is not chosen. If A, B is chosen, then we reverse the valuations, so that 1 is assigned when their measurements disagree and -1 when the agree. The CHSH inequality concerns the expectation of such a valuation. To formalize this, let $E\left(A^{(\prime)}, B^{(\prime)}\right)$ be the probability that the outcomes agree, given that $A^{(\prime)}, B^{(\prime)}$ is measured. Then we have a value

$$S = -E(A, B) + E(A', B) + E(A, B') + E(A', B')$$

The CHSH inequality states that under classical assumptions

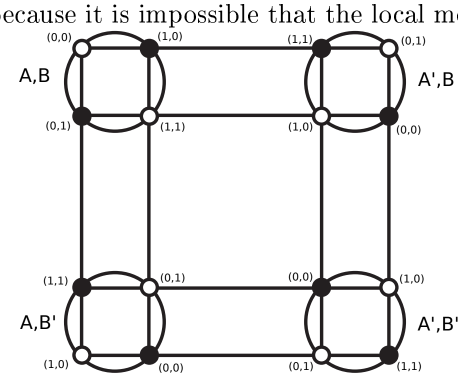
$$S \leq 2$$

We will see where this bound comes from. Yet using quantum mechanics, we may achive a value of

$$S = 2\sqrt{2}$$

Concretely, the measurements which realize this bound are: Alice may either measure in the standard basis (A), or in the basis $\left\{\frac{|0\rangle+|1\rangle}{\sqrt{2}}, \frac{|0\rangle-|1\rangle}{\sqrt{2}}\right\}$ (A'). Bob may either measure in the basis $\left\{\frac{\cos\left(\frac{\pi}{8}\right)|0\rangle-\sin\left(\frac{\pi}{8}\right)|1\rangle}{\sqrt{2}}, \frac{\sin\left(\frac{\pi}{8}\right)|0\rangle+\cos\left(\frac{\pi}{8}\right)|1\rangle}{\sqrt{2}}\right\}$($B$) and in the basis $\left\{\frac{\cos\left(\frac{\pi}{8}\right)|0\rangle+\sin\left(\frac{\pi}{8}\right)|1\rangle}{\sqrt{2}}, \frac{-\sin\left(\frac{\pi}{8}\right)|0\rangle+\cos\left(\frac{\pi}{8}\right)|1\rangle}{\sqrt{2}}\right\}$ (B'). (!) (TODO: check that this is accurate.)

To derive these bounds we first draw the exclusion graph for this scenario. The verticies of the graph are the 16 total outcomes which may occur (or, if you like, pairs of measurements and outcomes). Edges are drawn between vertices which cannot cooccur. For example, there is an edge between $(0, 0 \mid AB)$ and $(1, 0 \mid AB')$ because it is impossible that the local measurement A gives both 1 and 0.



Once the exclusion graph, H, is drawn, we obtain a weighting on the vertices which is determined by our valuation. For example, whenever the outcome $(1, 1 \mid AB')$ occurs, we add 1 to our counter, and whenever $(0, 0 \mid AB)$ occurs, we subtract 1 from our counter. Thus, each vertex gets a weight $w : V(G) \to \mathbb{R}$. We can insist that these weights be positive by adding 1 to all the experiments so that our final weighting corresponds to giving $w(v) = 2$ for the black vertices and $w(v) = 0$ for the white vertices. Typically, and in this case, all of the weights will be the same, except for those which are 0. Our weighting therefore identifies a subgraph, H'

To derive the classical bounds, observe that any classical state of the system must assign outcomes to all measurements, even those which have not occured. This means that a classical state will correspond to an independent set in the graph. It is intuitive (and easy to show, since S is a linear functional) S is maximized at a particular classical state rather than a mixture of them. The maximum value for S, classically, will correspond to an independent set and this independent set will restrict to an independent set of H'. Since $\alpha(H') = 3$, we can recover the classical bound on S by undoing our manipulations to the weighting.

$$S \leq_{classical} 2 \cdot 3 - 4 = 2$$

With respect to quantum mechanics, each of our outcomes corresponds to an eigenvector of an observable. Thus, each vertex in H' receives a vector. If the two outcomes associated with these vectors are exclusive, then the vectors must be orthogonal. Otherwise, it would be possible to measure the system, obtain one outcome, then measure the system again and obtain the other, exclusive outcome.

Therefore, the eigenvectors of the observables form an orthogonal representation of $\overline{H'}$. Recall our dual definition for $\vartheta(G) = \max_{\overline{OR}} \sum_i c(v_i)$ is exactly the maximum expected value in the quantum setting. Thus,

$$S \leq 2 \cdot \vartheta(G) - 4 = 2 \cdot \left(2 + \sqrt{2}\right) - 4 = 2\sqrt{2}$$

In this case, the bound is tight. This is not always the case, because the dimension of the optimal orthonormal representation may be too large.

(! Igonre this !)Interpretation

We can think of the verticies as properties that some system might posess, and the edges as the requirement that no system can possess both of those properties simultaneously. For example, (classically) an object cannot be entirely red and entirely green, and (quantum) an electron cannot be both spin up and spin down. Thus, the properties which this system actually posesses form an independent set in the graph.

The Kochen-Specker Theorem shows that it there is no coherent way assign values to measurements which have not occured. By doing this, we are assuming a fixed reality which our measurements reveal. An ontological state, classically, will correspond to an independent set in the graph.

In quantum mechanics, the measurements are projections (they can always be made so) and the state is a vector. Probabilities correspond to squared-inner product between the state and measurement. Incompatible properties will correpsond to orthogonal vectors. Observe that an orthonormal representation of the complement of the graph corresponds to such a scenario, and that the cost of each vertex is its probability.

Note that any collection of pairwise orthonormal measurement vectors can be grouped and we can find a self-adjoint matrix, an observable, whose eigenvectors include those measurement vectors. Thus, we choose a clique cover for the graph, and, for each clique, find an observable whose eigenvectors include the vectors from the clique. These observables will have other eigenvectors, which represent properties that does not concern our experiment. Due to these extra, ignored outcomes for our observables, the sum of the probabilities (costs) within a clique may be strictly less than 1.

The expected total number of properties (the sum of the probabilities) is bounded by α, classically. Due to the dual definition of the ϑ, ϑ bounds it with respect to quantum mechanics, and it's clear that this bound is tight. Thus, the discrepancy between ϑ and α represents a Bell inequality.

From Orthonormal Representations to Quantum Circuits

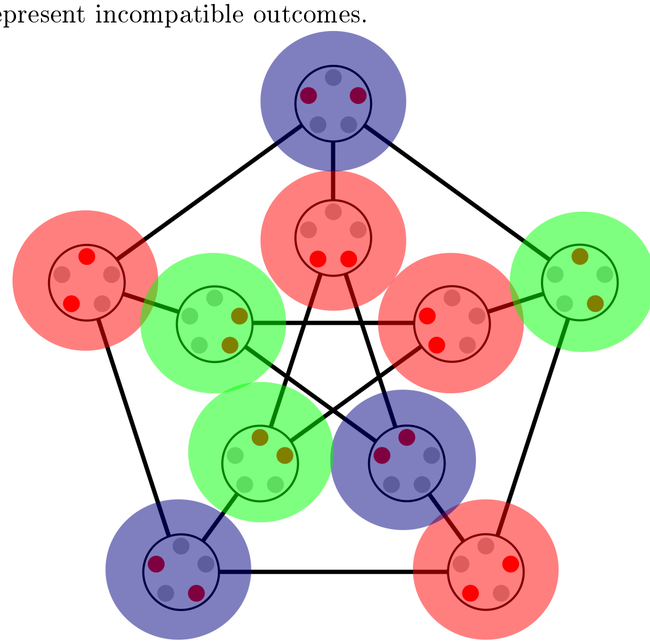
Next, we would like to construct quantum circuits which exhibit contextuality from given orthonormal representations. The procedure is as follows:

1. Choose a coloring of the graph
2. Produce an optimal orthonormal representation for this graph
3. Each color corresponds to a collection of pairwise orthonormal vectors. Extend each collection to an orthonormal basis.
4. Find the unitary transformations from the given bases to the standard (computational) basis.
5. All unitary transformations can be implemented by quantum gates.

(6) Our circuits consist of these unitary transformations, followed by measurement in the computational basis.

For a quantum circuit which takes some input state, ψ, applies a unitary transformation, U, then measures in the computational basis, the possible outcomes will be the elements of the computational basis. If x is one such basis element, the probability of measuring x is $\langle U\psi, x\rangle^2 = \langle \psi, U^{\star}x\rangle^2$. It is easier to think of the unitary transformations as moving the computational basis than moving the state. We will think of these quantum circuits as collections of bases under which to measure a particular state.

Since any two orthogonal vectors can be extended to a basis and we can perform a measurement (which will reveal one outcome) in that basis, two orthogonal vectors represent incompatible outcomes.



Our red vectors are: