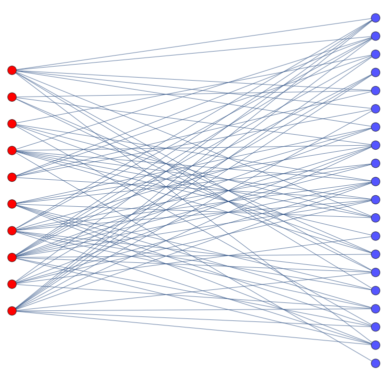
Graphs and Physics

K. Sutner
Carnegie Mellon University
Spring 2021

Movie Reviews



Suppose you have a bipartite graph that associates movies with viewers. Each edge has a weight representing the score given to the movie by a viewer.

Random Walks

How should one use this graph to give recommendations to a potential viewer?

The problem is to compute a score $\mathcal{S}(v,m)$ that reflects the expected score assigned by viewer v to movie m.

It is not hard to come up with some hacks, but it is difficult to get good scores in a computationally efficient manner. One idea that works fairly well is to to let

$$\mathcal{S}(v,m) = \text{hit}(v,m) + \text{hit}(m,v)$$

where $\text{hit}(x,y)$ indicates the expected length of a random walk from x to y (biased by edge weights).

The question now becomes: how do we compute these expected lengths?

Answer: Think electricity.

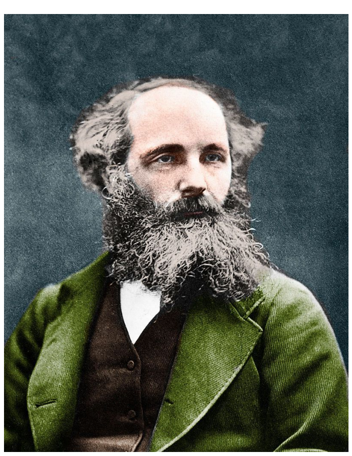
Preamble: James Clerk Maxwell

Lecturing in Cambridge, Maxwell sees a student dozing off, awakens him and the following dialogue ensues:

"Young man, what is electricity?"

"I'm terribly sorry, sir, I knew the answer but I have forgotten it."

"Gentlemen, you have just witnessed the greatest tragedy in the history of science. The one person who knew what electricity is has forgotten it."



Graph Algorithms

A search for "graph algorithms" produces 146 million hits on Google. The beginnings were a bit whimsical—recall Euler and his bridges—but today they are industrial strength. Ironically, Google's search algorithm is graph based.

How does one develop a graph algorithm? Graphs are part of classical combinatorics, easily formalizable in some nice elementary axiom system like Peano Arithmetic (at least the finite ones). So one would expect the standard machinery of algorithm development to come into play.

This works fine for reasonably simple graph properties, and for reasonably small graphs. But when faced with complicated properties on huge graphs, a little help from other sources is most welcome.

Physics to the Rescue

As it turns out, in some cases the inspiration comes from an entirely unexpected direction: physics.

This is surprising, since physics seems to have little to do with combinatorics, and is certainly not formalizable in Peano arithmetic.

Yet, in some sense, this is old news: in the 19th century, math and physics were closely coupled, in a mutually beneficial way (never mind Einstein's general relativity). But we are here talking about motivating ideas in the computational universe, a very, very different ballgame. Still, the links are surprisingly tight.

As it turns out, physics-based algorithms are very important in the analysis of large graphs, and are used for example in Google's local page rank algorithm (it's proprietary, so this is an educated guess).

Great, But ...

There is a vexing issue, though:

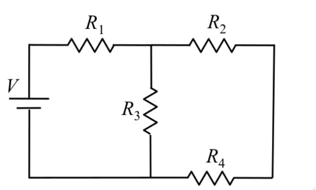
Question: How do we know these algorithms are correct?

Appeals to physical intuition are priceless to develop ideas and construct plausibility arguments, but it's a long way from there to a proof (think theorem prover). Axiomatizing all of physics is currently a pipedream.

But, we don't need all of physics, just a few pieces that need to be converted into a solid axiomatic framework. Beyond that, a little linear algebra is all we need, no big guns required.

Electric Circuits

We will consider circuits containing only resistors and just a single battery, typically expressed in diagrams like the following:

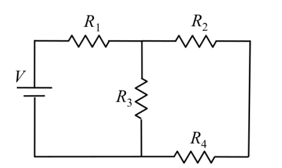


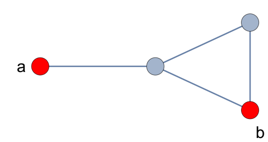
The voltage difference generated by the battery on the left will cause a current to flow. The strength of the current depends only on the voltage, the resistors and the layout of the circuit, and is uniquely determined.

Physical Axioms: We need rules such as Ohm's Law and Kirchhoff's Law allow us to compute the current.

Translation

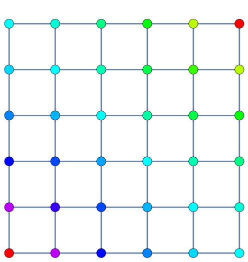
The first step is to translate the circuit into a two-terminal graph, an undirected graph $G = \langle \mathsf{V}, \mathsf{E} \rangle$ with two distinguished vertices.





More generally, we are dealing with graphs with a boundary and interior nodes. In general, the boundary could involve an arbitrary number of nodes. We will only consider boundaries with two nodes a and b (just one battery).

Current Driven Circuits



One question we would like to answer is the following:

What voltages are required at the nodes to cause a current of 1 unit to flow from a to b?

This gets fairly messy even for a simple graph like a grid.

Our intuition tells us that there has to be a unique solution, and that solution must obey certain symmetry principles. We should be able to derive these solutions formally.

Circuits and Graphs

We will think of the edges as being wires cum resistors in an electric circuit.

- Each edge carries a resistance R, equivalently a conductance $C = 1/R$.
- A voltage differential across an edge produces a current I.

We are interested in a potential $\boldsymbol{v} : \mathsf{V} \to \mathbb{R}$ that defines the voltage drops along an edge xy: $\boldsymbol{v}_x - \boldsymbol{v}_y$.

As is customary, we think about current flowing from a higher voltage to a lower one. The strength of the current depends only on the potential difference, though.

Omh's Law

The fundamental connection between current, voltage and resistance is:

Theorem (Ohm's Law, 1827)

$I = V/R = CV$

As a matter of physics, this can be established from measurements (or perhaps derived within the framework of some deeper theory, see Ohm-Maxwell for some arguments about this).

For us, this is an axiom, and can be used as such in proofs. Note, though, that this is quite different from clean equational logic, there are lots of unmentioned side conditions.

Voltage and Current

Here are the two basic questions that concern us:

Voltage Driven Suppose we have a potential that assigns voltages to each node (a voltage driven circuit). What is the resulting current?

Current Driven Suppose we force a particular flow through the circuit. What are the corresponding voltages?

We are mostly interested in the two-terminal case: say, we apply a voltage to a and b, and let the circuit settle down in some stable state. Then we measure the current from a to b.

Physical intuition tells us that the answers should be uniquely determined in either case, and closely related.

Simplify

Convention: For simplicity, we will assume all edges have resistance $R = 1$. Also, we will assume that the graph G is connected.

The general case is messier, but not substantially different.

We need enough axiomatic assumptions so we can formally prove that the answers to a basic questions are uniquely determined, and to derive a way to compute them. Clearly, Ohm's law is not enough, we need to understand how current flows through our network and how voltages behave.

In a sense, we would like to find an "equivalent" network

carrying the right resistance R and producing the same current.

Who Cares?

Recall our original problem, the movie recommendations?

We can think of two viewers as having similar taste in movies whenever the resistance between them is low (or the conductance is high). If a liked a movie a lot, we should recommend it to b. As a practical matter, this works out rather well. We can only go part of the way here, explaining some of the basic ideas.

If you want to know all the details, take 15-451 with Gary Miller.

Kirchhoff's Laws

Write $\boldsymbol{i}_{xy}$ for the current from node x to y (where $\boldsymbol{i}_{yx} = -\boldsymbol{i}_{xy}$). By $\Delta\boldsymbol{v}_e$ we mean the voltage drop across the edge e.

Theorem (Kirchhoff's Laws, 1845)

- *For all nodes* $x \neq a, b$: $\sum_{xy \in \mathsf{E}} \boldsymbol{i}_{xy} = 0$.
- *For any loop* $e_1, \ldots, e_k$: $\sum \Delta v_{e_i} = 0$.

These are referred to as the node law and the loop law. Again, for physics these are measurable observations, for us they are axioms. We will mostly require the node law.

Formally, $\boldsymbol{i} : \mathsf{E} \to \mathbb{R}$.

Consequences

Suppose $x \neq a, b$ is a node of degree d. As a direct consequence of Kirchhoff's law, the voltage at x is the (weighted) average of its neighbors:

$$\boldsymbol{v}_x = 1/d \sum_{xy \in \mathsf{E}} \boldsymbol{v}_y$$

Proposition

Suppose $\boldsymbol{v}_a = 0$ *and* $\boldsymbol{v}_b = v \geq 0$. *Then* $0 \leq \boldsymbol{v}_x \leq v$.

Proof. Assume $\boldsymbol{v}_x = u > v$ for some x. Let u be maximal such. Then all neighbors of x must also have voltage u. By induction, the whole connected component of x has voltage u, contradiction. Lower bound is similar. □

Harmonic Functions

We can generalize the idea that a node is "the average of its neighbors." Assume that G is connected. A map $f : \mathsf{V} \to \mathbb{R}$ is harmonic if

$$\forall x \neq a, b \left(f(x) = \frac{1}{\deg(x)} \sum\nolimits_{xy \in \mathsf{E}} f(y) \right)$$

Lemma

Let f and g harmonic.

- *Suppose $f(a) \leq f(b)$. Then $f(a) \leq f(x) \leq f(b)$.*
- *Let $f(a) = g(a)$ and $f(b) = g(b)$. Then $f = g$.*

We have seen the argument for the bounds. For uniqueness note that $f - g$ is also harmonic.

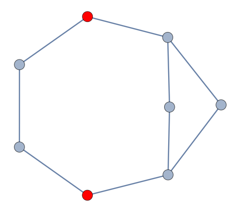
This also works for more general boundaries $B \subseteq \mathsf{V}$.

Digression: Series-Parallel Graphs

Consider two-terminal graphs with distinguished nodes s and t. Suppose we have two such graphs, vertex disjoint.

- Series Composition: union of G_1 and G_2, then merge t_1 and s_2.
- Parallel Composition: union of G_1 and G_2, then merge s_1 and s_2, and t_1 and t_2.

A graph is series-parallel if it can be generated from K_2 by a sequence of series/parallel compositions.



SP is Easy

Theorem (Tarjan, Lawler 1982)

One can check in linear time whether a graph is series-parallel. Moreover, one can compute its series-parallel decomposition in linear time.

Corollary

We can compute the current from s to t in a series-parallel graph in linear time.

For the last example, a unit voltage drop will produce a current of $2/3$.

This works, since series-parallel graphs have a simple inductive structure. Alas, general graphs require more effort.

First, some notions from graph theory.

Graph Matrices

For simplicity, think of the following matrices as being of type $\mathbb{R}^{n \times n}$ (actually, they are integer valued).

- A adjacency matrix,
- D degree matrix,
- Q Laplacian or Kirchhoff matrix: $\mathsf{Q} = \mathsf{D} - \mathsf{A}$.

Lemma

The Laplacian has 0 row-sums, is symmetric and positive semidefinite.

It has rank $n - 1$ and its nullspace is spanned by $\mathbf{1}$.

Its eigenvalues are real $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.

The eigenvalues are not divorced from the graph: the multiplicity of λ_1 is the number of connected components of G. λ_2 is the algebraic connectivity; it is positive iff G is connected.

Proof Lemma

0 row-sums and symmetry are obvious.

For positive semidefiniteness let $\mathsf{E}^{\to}$ be any orientation on the edges and note that

$$\boldsymbol{x}^T \mathsf{Q}\, \boldsymbol{x} = \sum_{ij \in \mathsf{E}^{\to}} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2$$

It follows, for example, that Q can be written as $M^T M$, $M \in \mathbb{R}^{n \times n}$. In fact, M can even be chosen to be upper triangular (Cholesky decomposition).

Moreover, the null space of Q is $\mathsf{span}(\mathbf{1})$ and Q has rank $n - 1$.

The eigenvalue claim is a standard result from linear algebra: every symmetric (actually, Hermitian) matrix admits an orthonormal basis of eigenvectors, and the eigenvalues are real.

□

Solving Linear Systems with Q

Let $\mathfrak{N}$ and $\mathfrak{C}$ denote the null space and column space of a matrix. A standard fact of linear algebra is that for $M \in \mathbb{R}^{n \times n}$ is

$$\mathfrak{N}(M^T) = \mathfrak{C}(M)^{\perp}$$

Since Q is symmetric we get

$$\mathfrak{C}(\mathsf{Q}) = \mathfrak{N}(\mathsf{Q}^T)^{\perp} = \mathfrak{N}(\mathsf{Q})^{\perp} = \mathsf{span}(\mathbf{1})^{\perp} = \{\, \boldsymbol{x} \mid \sum x_i = 0 \,\}$$

So we can solve a linear system

$$\mathsf{Q}\, \boldsymbol{v} = \boldsymbol{b}$$

as long as $\sum \boldsymbol{b}_i = 0$. All solutions then look like $\boldsymbol{v} + c\,\mathbf{1}$.

This is a good sign: we are interested in potential differences, not the actual values.

So What?

The physics interpretation of $\mathsf{Q}\,\boldsymbol{v}$ is

$$\mathsf{Q}\,\boldsymbol{v} = \text{ vector of local currents at each node}$$

These are the currents we need to inject to maintain the voltage.

Cheap example:

$$\mathsf{Q}\,\boldsymbol{v} = \begin{pmatrix} v_1 - v_2 \\ 2v_2 - v_1 - v_3 \\ 2v_3 - v_2 - v_4 \\ v_4 - v_3 \end{pmatrix}$$

Demand

Let us return to our old scenario: we apply voltages at a and b, and let the circuit determine the rest. Say, we are interested in the case where we get a unit current: $\boldsymbol{i}_{ab} = 1$.

Write $\boldsymbol{e}_{ab} = \boldsymbol{e}_a - \boldsymbol{e}_b$ for the so-called demand vector. So we need to solve

$$\mathsf{Q}\,\boldsymbol{v} = \boldsymbol{e}_{ab}$$

But this works by our previous characterization of the column space of Q.

Moreover, we can assume $\boldsymbol{v}_a = 0$ or $\boldsymbol{v}_b = 0$ if that's convenient.

Effective Resistance

Consider a flow associated with potentials $\boldsymbol{v}$ as above. The effective resistance of the circuit between a and b is defined as

$$R_{ab} = (\boldsymbol{v}_a - \boldsymbol{v}_b)/\boldsymbol{i}_{ab}$$

Theorem

Effective resistance is well-defined.

Again, this seems totally obvious from a physics perspective, but required an argument in our axiomatic setting: we have to prove that our axioms imply that there is a unique solution for the unit flow from a to b, depending only on G.

Note that the solution $\boldsymbol{v}$ from above is not unique, but adding a constant to the potential does not affect the current nor the voltage drop.

Alternatively . . .

We could also clamp the voltages, say, $\boldsymbol{v}_a = 1$, $\boldsymbol{v}_b = 0$ and then determine the induced current by solving

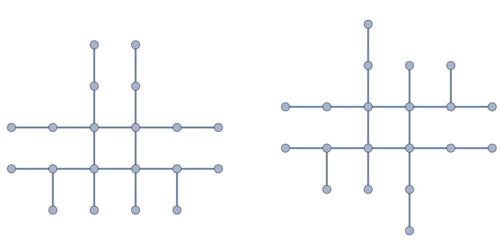
$$\mathsf{Q} \cdot (1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{n-1}, 0) = (I, 0, \ldots, 0, -I)$$

But note that this is not the form a standard linear solver requires: we need something like $M\boldsymbol{x} = \boldsymbol{b}$. But here we are dealing with a boundary value problem.

Resistance Equivalence

Similarly, the effective graph resistance of G is $R_G = \sum_{a \neq b} R_{ab}$.

Suppose G and H are two n-node graphs. They are called resistance equivalent if $R_G = R_H$. Perhaps surprisingly, it is not so easy to find non-isomorphic graphs that are equivalent in this sense. The smallest ones have 9 points. Here is a 20 point example with resistance 561.



Note the similarity—whatever that means.

Incidence Matrix

Define an V × E matrix, the so-called (signed) incidence matrix of G, as follows. Choose any orientation of the edges of G and set

$$\mathsf{B}(v,e) = \begin{cases} -1 & \text{if } v \text{ is the source of } e, \\ 1 & \text{if } v \text{ is the target of } e, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\boldsymbol{v}$ be a potential on the vertices, the $\mathsf{B}^T \boldsymbol{v}$ is the voltage differential at each edge.

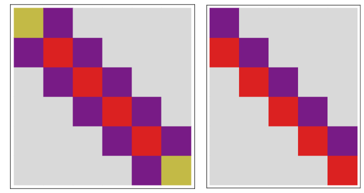
For us, that is the same as the current flow (though in general we would have to multiply by conductance). Correspondingly, $\mathsf{B}\,\mathsf{B}^T \boldsymbol{v}$ is the residual current for each edge.

Proposition

$\mathsf{Q} = \mathsf{B}\,\mathsf{B}^T$.

Paths

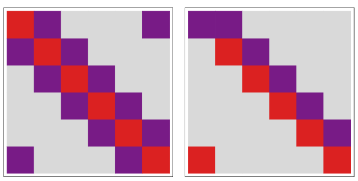
P_6, matrices Q and B^T:



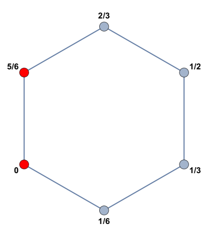
Potentials:

Cycles

C_6, matrices Q and B^T:

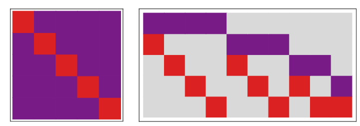


Potentials:

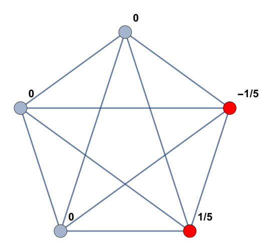


Complete Graphs

K_6, matrices Q and B^T:

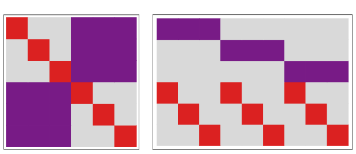


Potentials:

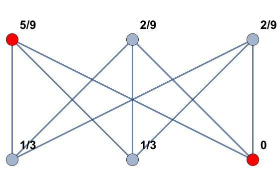


Complete Bipartite

$K_{3,3}$, matrices Q and B^T:

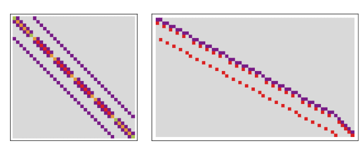


Potentials:

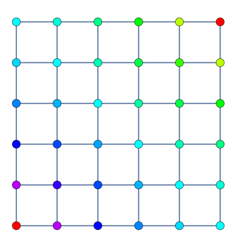


Complete Bipartite

$G_{6,6}$, matrices Q and B^T:



Potentials:



Flows

So far we have only looked at current flows that conform to Kirchhoff's law. Mathematically, it is natural to generalize to arbitrary flows, temporarily ignoring physics.

Let $\mathsf{E}^{\rightarrow}$ be an orientation on the edges. A general flow is a map $f : \mathsf{E}^{\rightarrow} \to \mathbb{R}$. The space $\mathcal{F}$ of all flows has dimension m.

A potential flow is one that is induced by a potential:

$$\mathcal{P}_G = \{ \mathsf{B}^T \boldsymbol{v} \mid \boldsymbol{v} \in \mathbb{R}^n \}$$

A circulation flow just moves stuff around in circles:

$$\mathcal{C}_G = \{ f \in \mathbb{R}^m \mid B\, f = 0 \}$$

As usual, define $f_x = \sum_{xy \in \mathsf{E}^{\rightarrow}} f(xy)$.

f is a unit flow from a to b if $-f_b = f_a = 1$ but otherwise $f_x = 0$.

Flow Decomposition

Theorem

Suppose G is connected.
Then $\mathcal{F} = \mathcal{P} \oplus \mathcal{C}$ and $\mathcal{P}$ has dimension $n-1$.

Proof.

Consider a spanning tree T of G, and let $E_0 = \mathsf{E} - T$ be the non-tree edges. So $|E_0| = m - n + 1$.

Any flow on E_0 can be extended to a circulation flow. On the other hand, any circulation flow is already determined by the subflow on E_0. Thus E_0 provides a basis for $\mathcal{C}$.

Orthogonality follows directly from the definitions.

□

Dirichlet's Minimum Principle

From a physics perspective, potential flows are distinguished in another way:

$$D = \boldsymbol{v}^T \mathsf{Q}\, \boldsymbol{v} = \sum_{ij \in \mathsf{E}^{\rightarrow}} (\boldsymbol{v}_i - \boldsymbol{v}_j)^2$$

is the energy dissipation associated with $\boldsymbol{v}$.

Dirichlet's principle asserts that this dissipation is minimized by the potential flows we considered in the last section: we want a potential $\boldsymbol{v}$ that produces a unit current, clamping the voltages $\boldsymbol{v}_a$ and $\boldsymbol{v}_b$. The circuit takes care of the rest, and it does so by minimizing energy dissipation.

Application

Question: What happens if we remove one edge uv from G?

Intuitively, it is entirely clear that dissipation can only decrease. Here is a little Gedankenexperiment based on Dirichlet's principle.

First, clamp the voltages at a, b, u and v. Then dissipation goes down by the amount at uv, the rest of the circuit is oblivious.

Now release the voltage at u and v (but keep a and b clamped). The circuit adjusts and finds a potentially even lower dissipation by Dirichlet.

So removing an edge can only drive dissipation down. But note that, depending on the edge chosen, nothing may change.

Thomson's Principle

Here is another minimum principle, from the current driven perspective.

Theorem

Let f and g be unit flows from a to b, f a potential flow. Then $\|f\| \leq \|g\|$.

Proof.

Decompose g: $g = g_p + g_c = f + g_c$ since there is only one unit potential flow from a to b. But then

$$\begin{aligned}\langle g, g \rangle &= \langle f + g_c, f + g_c \rangle \\ &= \langle f, f \rangle + \langle g_c, g_c \rangle + 2\langle f, g_c \rangle \\ &= \langle f, f \rangle + \langle g_c, g_c \rangle \\ &\geq \langle f, f \rangle\end{aligned}$$

□

Rayleigh's Law

As usual, fix $V = \boldsymbol{v}_a - \boldsymbol{v}_b$, producing current $I = \boldsymbol{i}_{ab}$. Let $R = R_{ab}$ be the effective resistance. Then $I = V/R$ and one can check that

$$D = IV = V^2/R = RI^2$$

But then the following is an immediate consequence of Dirichlet's principle:

Theorem

Removing an edge from G can only increase the effective resistance.

One can show that resistance strictly increases when the removed edge is not "superfluous."

More Formally

Proof.

Let G' be a graph obtained from G by removing an edge (more generally, we could increase resistance anywhere).

Note that $R_{ab} = \|f\|^2$ for the unit potential flow f from a to b. Define g similarly for G'. But then

$$\begin{aligned} R'_{ab} &= \sum_{e \in \mathsf{E}'} g_e^2 \\ &\geq \sum_{e \in \mathsf{E}} g_e^2 \\ &\geq \sum_{e \in \mathsf{E}} f_e^2 = R_{ab} \end{aligned}$$

The second inequality follows from Thomson's principle.

□

Where Are We?

We can express basic principles of circuit physics in terms of linear algebra.

In particular, Ohm's law looks like so

$$\boldsymbol{i} = B^T \boldsymbol{v}$$

and Kirchhoff's conservation law like this

$$B\,\boldsymbol{i} = \boldsymbol{e}_{ab}$$

or, in terms of the Laplacian,

$$\mathsf{Q}\,\boldsymbol{v} = B\,B^T \boldsymbol{v} = \boldsymbol{e}_{ab}$$

Can we get computational mileage out of this?

Eigendecomposition

Lemma

Let A be an $n \times n$ matrix with n distinct eigenvectors. Then

$$A = U\,\Lambda\,U^{-1}$$

where Λ is the diagonal matrix of the eigenvalues, and U is the matrix whose columns are the eigenvectors.

If A is in addition symmetric, we can choose the eigenvectors, say $\boldsymbol{q}_i$, to be orthonormal:

$$\|\boldsymbol{q}_i\| = 1 \qquad \boldsymbol{q}_i \perp \boldsymbol{q}_j \text{ for } i \neq j$$

But then U is orthogonal, so $U^{-1} = U^T$ and we can avoid matrix inversion.

Clearly this all applies to the Laplacian Q.

Moore-Penrose Pseudo-Inverse

For a matrix M to have an inverse M^{-1} is a rather special property.

Computationally, it sometimes suffices to have a matrix $\widehat{M}$ that behaves more or less like an inverse, This pseudo-inverse exists even though M may not be invertible, and not even a square matrix.

In essence, we would like to have

$$\begin{gathered} M\,\widehat{M}\,M = M \\ \widehat{M}\,M\,\widehat{M} = \widehat{M} \\ \widehat{M}\,M,\ M\,\widehat{M} \text{ symmetric} \end{gathered}$$

Laplacian Pseudo-Inverse

Recall that for Q we have $\lambda_1 = 0$ and $0 < \lambda_i$ for $i \geq 2$. Using the orthonormal basis from above, we have

$$\mathsf{Q} = U\,\Lambda\,U^T = \sum_{i=1}^{n} \lambda_i \boldsymbol{q}_i \boldsymbol{q}_i^T$$

Now remove the 0 eigenvalue and let

$$\widehat{\mathsf{Q}} = \sum_{i=2}^{n} \lambda_i \boldsymbol{q}_i \boldsymbol{q}_i^T$$

Claim: $\widehat{\mathsf{Q}}$ is the pseudo-inverse of Q.

This is easy to check using orthonormality of U.

Computing Effective Resistance

The Laplacian Q is singular and symmetric, its kernel is span$(\mathbf{1})$. Let $\widehat{\mathsf{Q}}$ be the pseudo-inverse as above. Then $\widehat{\mathsf{Q}}$ is

- the inverse of Q on span$(\mathbf{1})^{\perp}$, and
- zero on span$(\mathbf{1})$.

Instead of solving a linear system, we can compute effective resistance by matrix-vector multiplication (assuming, of course, we have $\widehat{\mathsf{Q}}$).

So since $\boldsymbol{e}_{ab}$ is orthogonal to the kernel, $\mathsf{Q}\,\boldsymbol{v} = \boldsymbol{e}_{ab}$ translates into

$$\boldsymbol{v} = \widehat{\mathsf{Q}}\,\boldsymbol{e}_{ab}$$

$$R_{ab} = \boldsymbol{e}_{ab}^T\,\widehat{\mathsf{Q}}\,\boldsymbol{e}_{ab}$$

Total Resistance

Recall that the effective graph resistance of G was defined to be $R_G = \sum_{a \neq b} R_{ab}$. From the last slide, this is fairly easy to compute once we have $\widehat{\mathsf{Q}}$.

But we can even do better, much better:

Theorem

$R_G = n \sum_{i \geq 2} \lambda_i^{-1}$.

We'll skip the proof.

This approach is also useful for bounds. For example,

$$n/\lambda_2 < R_G \leq n(n-1)/\lambda_2$$