Graphs and Networks Lecture 14

Monotonicity and its Failures

Daniel A. Spielman October 15, 2013

14.1 Disclaimer

These notes are not necessarily an accurate representation of what happened in class. They are a combination of what I intended to say with what I think I said. They have not been carefully edited.

14.2 Overview

14.3 Effective Spring Constants

Consider a spring network. As in last lecture, we model it by a weighted graph $G = (V, E, w)$, where $w_{a,b}$ is the spring constant of the edge (a, b). Recall that a stronger spring constant results in a stronger connection between a and b.

Now, let s and t be arbitrary vertices in V. We can view the network as a large, complex spring connecting s to t. We then ask for the spring constant of this complex spring. We call it the *effective spring constant* between s and t.

To determine what it is, we recall the definition of the spring constant for an ordinary spring: the potential energy in a spring connecting a to b is the spring constant times times the square of the length of the spring, divided by 2. We use this definition to determine the effective spring constant between s and t.

Recall again that if we fix the positions of s and t on the real line, say to 0 and 1, then the positions $\boldsymbol{x}$ of the other vertices will minimize the total energy:

$$\mathcal{E}\left(\boldsymbol{x}\right) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{(a,b) \in E} w_{a,b} (\boldsymbol{x}(a) - \boldsymbol{x}(b))^2. \tag{14.1}$$

As s and t are separated by a distance of 1, we may define twice this quantity to be the effective spring constant of the entire network between s and t. To verify that this definition is consistent, we should consider what happens if the displacement between s and t is something other than 1. If we fix the position of s to 0 and the position of t to y, then the homogeniety of the expression for energy (14.1) tells us that the vector $y\boldsymbol{x}$ will minimize the energy subject to the boundary conditions. Moreover, the energy in this case will be $y^2/2$ times the effective spring constant.

14.4 Monotonicity

Rayleigh's Monotonicity Principle tells us that if we alter the spring network by decreasing some of the spring constants, then the effective resistance between s and t will not increase.

Theorem 14.4.1. *Let $G = (V, E, w)$ be a weighted graph and let $\hat{G} = (V, E, \hat{w})$ be another weighted graph with the same edges and such that*

$$\hat{w}_{a,b} \leq w_{a,b}$$

for all $(a, b) \in E$. For vertices s and t, let $c_{s,t}$ be the effective spring constant between s and t in G and let $\hat{c}_{s,t}$ be the analogous quantity in $\hat{G}$. Then,

$$\hat{c}_{s,t} \leq c_{s,t}.$$

Proof. Let $\boldsymbol{x}$ be the vector of minimum energy in G such that $\boldsymbol{x}(s) = 0$ and $\boldsymbol{x}(t) = 1$. Then, the energy of $\boldsymbol{x}$ in $\hat{G}$ is no greater:

$$\frac{1}{2} \sum_{(a,b) \in E} \hat{w}_{a,b} (\boldsymbol{x}(a) - \boldsymbol{x}(b))^2 \leq \frac{1}{2} \sum_{(a,b) \in E} w_{a,b} (\boldsymbol{x}(a) - \boldsymbol{x}(b))^2 = c_{s,t}.$$

So, the minimum energy of a vector $\boldsymbol{x}$ in $\hat{G}$ such that $\boldsymbol{x}(s) = 0$ and $\boldsymbol{x}(t) = 1$ will be at most $c_{s,t}$, and so $\hat{c}_{s,t} \leq c_{s,t}$. □

While this principle seems very simple and intuitively obvious, it turns out to fail in just slightly more complicated situtations. Before we examine them, I will present the analogous material for electrical networks.

14.5 Effective Resistance

There are two (equivalent) ways to define the *effective resistance* between two vertices in a network of resistors. The first is to start with the formula

$$V = IR,$$

or, as I prefer to write it,

$$\boldsymbol{i}(a, b) = \frac{\boldsymbol{v}(a) - \boldsymbol{v}(b)}{r_{a,b}},$$

This formula tells us that if we have one resistor between a and b and we fix the voltage of a to 1 and the voltage of b to 0, then the amount of current that will flow from a to b is the reciprocal of the resistance. It also tells us that if we want to flow one unit of current, then we need to place a potential difference of $r_{a,b}$ between a and b. Recall that we define the weight of an edge to be the reciprocal of its resistance, as high resistance corresponds to poor connectivity. We can use this formula to define the effective resistance between two vertices s and t in an arbitrary complex network of resistors: we define the effective resistance between s and t to be the potential difference needed to flow one unit of current from s to t.

Algebraically, define $\boldsymbol{i}_{ext}$ to be the vector

$$\boldsymbol{i}_{ext}(a) = \begin{cases} 1 & \text{if } a = s \\ -1 & \text{if } a = t \\ 0 & \text{otherwise} \end{cases}.$$

This corresponds to a flow of 1 from s to t. We then solve for the voltages that realize this flow:

$$\boldsymbol{L}\boldsymbol{v} = \boldsymbol{i}_{ext},$$

by

$$\boldsymbol{v} = \boldsymbol{L}^{+}\boldsymbol{i}_{ext}.$$

We thus have

$$\boldsymbol{v}(s) - \boldsymbol{v}(t) = \boldsymbol{i}_{ext}^{T}\boldsymbol{v} = \boldsymbol{i}_{ext}^{T}\boldsymbol{L}^{+}\boldsymbol{i}_{ext} \stackrel{\text{def}}{=} \mathrm{R_{eff}}(s,t).$$

This agrees with the other natural approach to defining effective resistance: twice the energy dissipation when we flow one unit of current from s to t.

Theorem 14.5.1. *Let $\boldsymbol{i}$ be the electrical flow of one unit from vertex s to vertex t in a graph G. Then,*

$$\mathrm{R_{eff}}_{s,t} = \mathcal{E}\left(\boldsymbol{i}\right).$$

Proof. Recalling that $\boldsymbol{i}_{ext} = \boldsymbol{L}\boldsymbol{v}$, we have

$$\mathrm{R_{eff}}_{s,t} = \boldsymbol{i}_{ext}^{T}\boldsymbol{L}^{+}\boldsymbol{i}_{ext} = \boldsymbol{v}^{T}\boldsymbol{L}\boldsymbol{L}^{+}\boldsymbol{L}\boldsymbol{v} = \boldsymbol{v}^{T}\boldsymbol{L}\boldsymbol{v} = \mathcal{E}\left(\boldsymbol{v}\right).$$

□

Rayleigh's Monotonicity Theorem was originally stated for electrical networks.

Theorem 14.5.2 (Rayleigh's Monotonicity)**.** *The effective resistance between a pair of vertices cannot be decreased by increasing the resistance of some edges.*

14.6 Examples

In the case of a path graph with n vertices and edges of weight 1, the effective resistance between the extreme vertices is $n-1$.

In general, if a path consists of edges of resistance $r(1,2), \ldots, r(n-1,n)$ then the effective resistance between the extreme vertices is

$$r(1,2) + \cdots + r(n-1,n).$$

To see this, set the potential of vertex i to

$$\boldsymbol{v}(i) = r(i, i+1) + \cdots + r(n-1, n).$$

Ohm's law then tells us that the current flow over the edge $(i, i+1)$ will be

$$\left(\boldsymbol{v}(i) - \boldsymbol{v}(i+1)\right) / r(i, i+1) = 1.$$

If we have k parallel edges between two nodes s and t of resistances $r_1, \ldots, r_k$, then the effective resistance is

$$\mathrm{R_{eff}}(s, t) = \frac{1}{1/r_1 + \cdots + 1/r_k}.$$

Again, to see this, note that the flow over the ith edge will be

$$\frac{1/r_i}{1/r_1 + \cdots + 1/r_k},$$

so the total flow will be 1.

14.7 Breakdown of Monotonicity

We will now exhibit a breakdown of monotonicity in networks of nonlinear elements. In this case, we will consider a network of springs and wires. For examples in electrical networks with resistors and diodes or for networks of pipes with valves, see [PP03] and [CH91].

There will be 4 important vertices in the network that I will describe, a, b, c and d. Point a is fixed in place at the top of my aparatus. Point d is attached to an object of weight 1. The network has two springs of spring constant 1: one from point a to point b and one from point c to point d. There is a very short wire connecting point b to point c.

As each spring is supporting one unit of weight, each is stretched to length 1. So, the distance from point a to point d is 2.

I now add two more wires to the network. One connects point a to point c and the other connects point b to point d. Both have lengths $1 + \epsilon$, and so are slack. Thus, the addition of these wires does not change the position of the weight.

I now cut the small wire connecting point b to point c. While you would expect that removing material from the supporting structure would cause the weight to go down, it will in fact move up. To see why, let's analyze the resulting structure. It consists of two suppors in parallel. One consists of a spring from point a to point b followed by a wire of length $1 + \epsilon$ from point b to d. The other has a wire of length $1 + \epsilon$ from point a to point c followed by a spring from point c to point d. Each of these is supporting the weight, and so each carries half the weight. This means that the length of the springs will be $1/2$. So, the distance from a to d should be essentially $3/2$.

This sounds like a joke, but we will see in class that it is true. The measurements that we get will not be exactly 2 and $3/2$, but that is because it is difficult to find ideal springs at Home Depot.

In the example with resistors and diodes, one can increase electrical flow between two points by *cutting* a wire!

14.8 Traffic Networks

I will now explain some analogous behavior in traffic networks. We will examine the more formally in the next lecture.

We will use a very simple model of a road in a traffic network. It will be a directed edge between two vertices. The rate at which traffic can flow on a road will depend on how many cars are on the road: the more cars, the slower the traffic. I will assume that our roads are linear. That is, when a road has flow f, the time that it takes traffic to traverse the road is

$$af + b,$$

for some nonnegative constants a and b. I call this the characteristic function of the road.

We first consider an example of Pigou consisting of two roads between two vertices, s and t. The slow road will have characteristic function 1: think of a very wide super-highway that goes far out of the way. No matter how many cars are on it, the time from s to t will always be 1. The fast road is better: its characteristic is f. Now, assume that there is 1 unit of traffic that would like to go from s to t.

A global planner that could dictate the route that everyone takes could minimize the average time of the traffic going from s to t by assigning half of the traffic to take the fast road and half of the traffic to take the slow road. In this case, half of the traffic will take time 1 and half will take time 1/2, for an average travel time of 3/4. To see that this is optimal, let f be the fraction of traffic that takes the fast road. Then, the average travel time will be

$$f \cdot f + (1 - f) \cdot 1 = f^2 - f + 1.$$

Taking derivatives, we see that this is minimized when

$$2f - 1 = 0,$$

which is when $f = 1/2$.

On the other hand, this is not what people will naturally do if they have perfect information and freedom of choice. If a $f < 1$ fraction of the flow is going along the fast road, then those travelling on the fast road will get to t faster than those going on the slow road. So, anyone going on the slow road would rather take the fast road. So, all of the traffic will wind up on the fast road, and it will become not-so-fast. All of the traffic will take time 1.

We call this the Nash Optimal solution, because it is what everyone will do if they are only maximizing their own benefit. You should be concerned that this is not as well as they would do if they allowed some authority to dictate their routes. For example, the authority could dictate that half the cars go each way every-other day, or one way in the morning and another at night.

Let's see an even more disturbing example.

14.9 Braes's Paradox

We now examine Braes's Paradox, which is analogous to the troubling example we saw with springs and wires. This involves a network with 4 vertices, a, b, c, and d. All the traffic starts at $s = a$ and wants to go to $t = d$. There are slow roads from s to c and from d to t, and fast roads from s to d and from c to t. If half of the traffic goes through route sct and the other half goes through route sdt, then all the traffic will go from s to t in time $3/2$. Moroever, noone can improve their lot by taking a different route, so this is a Nash equilibrium.

We now consider what happens if some well-intentioned politician decides to build a very fast road connecting c to d. Let's say that its characteristic function is 0. This opens up a faster route: traffic can go from s to c to d to t. If no one else has changed route, then this traffic will reach t in 1 unit of time. Unfortunately, once everyone realizes this all the traffic will take this route, and everyone will now require 2 units of time to reach t.

Let's prove that formally. Let p_1, p_2 and p_3 be the fractions of traffic going over routes sct, sdt, and $scdt$, respectively. The cost of route sct is $p_1 + p_3 + 1$. The cost of route sdt is $p_2 + p_3 + 1$. And, the cost of route $scdt$ is $p_3 + p_3$. So, as long as p_3 is less than 1, the cheapest route will be $scdt$. So, all the traffic will go that way, and the cost of every route will be 2.

14.10 The Price of Anarchy

In any traffic network, we can measure the average amount of time it takes traffic to go from s to t under the optimal flow. We call this the cost of the social optimum, and denote it by $\mathrm{Opt}(G)$. When we let everyone pick the route that is best for themselves, the resulting solution is a Nash Equilibrium, and we denote it by $\mathrm{Nash}(G)$.

The "Price of Anarchy" is the cost to society of letting everyone do their own thing. That is, it is the ratio

$$\frac{\mathrm{Nash}(G)}{\mathrm{Opt}(G)}.$$

In these examples, the ratio was $4/3$. In the next lecture, we will show that the ratio is never more than $4/3$ when the cost functions are linear. If there is time today, I will begin a more formal analysis of $\mathrm{Opt}(G)$ and $\mathrm{Nash}(G)$ that we will need in our proof.

14.11 Nash optimum

Let the set of s-t paths be $P_1, \ldots, P_k$, and let α_i be the fraction of the traffic that flows on path P_i. In the Nash equilibrium, no car will go along a sub-optimal path. Assuming that each car has a negligible impact on the traffic flow, this means that every path P_i that has non-zero flow must have minimal cost. That is, for all i such that $\alpha_i > 0$ and all j

$$c(P_i) \leq c(P_j).$$

14.12 Social optimum

Society in general cares more about the average time its takes to get from s to t. If we have a flow that makes this average time low, everyone could rotate through all the routes and decrease the total time that they spend in traffic. So, the social cost of the flow f is

$$\begin{aligned} c(\alpha_1, \ldots, \alpha_k) &\stackrel{\text{def}}{=} \\ \sum_i \alpha_i c(P_i) &= \sum_i \alpha_i \sum_{e \in P_i} c_e(f_e) \\ &= \sum_e c_e(f_e) \sum_{i : e \in P_i} \alpha_i \\ &= \sum_e c_e(f_e) f_e. \end{aligned}$$

Theorem 14.12.1. *All local minima of the social cost function are global minima. Moreover, the set of global minima is convex.*

Proof. This becomes easy once we re-write the cost function as

$$\sum_e c_e(f_e) f_e = \sum_e a_e f_e^2 + b_e f_e$$

and recall that we assumed that a_e and b_e are both at least zero. The cost function on each edge is convex. It is strictly convex if $a_e > 0$, but that does not matter for this theorem.

If you take two flows, say f^0 and f^1, the line segments of flows between them contains the flows of the form f^t where

$$f_e^t = t f_e^1 + (1 - t) f_e^0,$$

for $0 \leq t \leq 1$.

By the convexity of each cost function, we know that the cost of any flow f^t is at most the maximum of the costs of f^0 and f^1. So, if f^1 is the global optimum and f^0 is any other flow with higher cost, the flow f^ϵ will have a social cost lower than f^0. This means that f^0 cannot be a local optimum. Similarly, if both f^0 and f^1 are global optima, then f^t must be as well. □

References

[CH91] Joel E Cohen and Paul Horowitz. Paradoxical behaviour of mechanical and electrical networks. 1991.

[PP03] Claude M Penchina and Leora J Penchina. The braess paradox in mechanical, traffic, and other networks. *American Journal of Physics*, 71:479, 2003.