Form versus Function : Walking the Line

Amrish Deshmukh, Niko Stahl, Tarun Chitra

November 18, 2008

Abstract

The Arts quad is a space that is subject to several competing interests. Its design is expected to offer the opportunity for an individual to move conveniently from one end to another, to keep general maintenance costs feasible and to allow for activities ranging from lectures to snowball fights.

This paper proposes a mathematical model, which provides a way of intelligently designing the path network of the quad. The centerpiece of our model is a cost function, which evaluates the feasibility of a given path configuration.

To explore the set of feasible path configuration we wrote an algorithm that randomly generates samples of this set. We then improved on this search by constructing an optimization algorithm inspired by Markov chain Monte Carlo methods. We believe this improved search has found a local minimum path configuration, as it appears stable under perturbation.

Contents

Part I
Problem Statement

The task is to redesign the Arts quad walkways using a mathematical model that will help us determine a preferred design. Beyond the general fact that minimizing the total length of the paths and maximizing the areas of contiguous lawns are preferable, we are asked to consider the following criteria:

- Path maintenance costs
- Landscaping costs
- Pedestrian traffic and behavior
- The creation of unofficial paths and its impact on the lawn
- The general appeal of the quad

To implement these criteria in our model, we are provided with the following principles:

- The path maintenance cost is proportionate to the total path length.
- The landscaping cost depends on the number of contiguous lawns, the creation of unofficial paths (as a result of pedestrians leaving the paved paths to arrive at their destination more quickly) and the geometry of contiguous lawn.
- If the path between two points is 15% longer than the straight line connecting the points, a pedestrian will leave the path and cut across the quad.
- An average pedestrian might leave the path if it implies saving more than 10% of the total length the path.

Part II
The Arts Quadrangle as a Graph

Graph theory has been an important tool in exploring problems which range from determining the neural network of nematode C. elegans to finding the cause of failure in electrical power grids[1]. By framing our walkway design problem in the language of graphs, we can readily extract the key relationships between structure and function.

We describe the **Arts quadrangle (hereby referred to as the Arts quad or simply quad) as a graph of 10 nodes**, which represent the most common points of entry and exit to the quad (see figure 1 below).

[1] See [1]

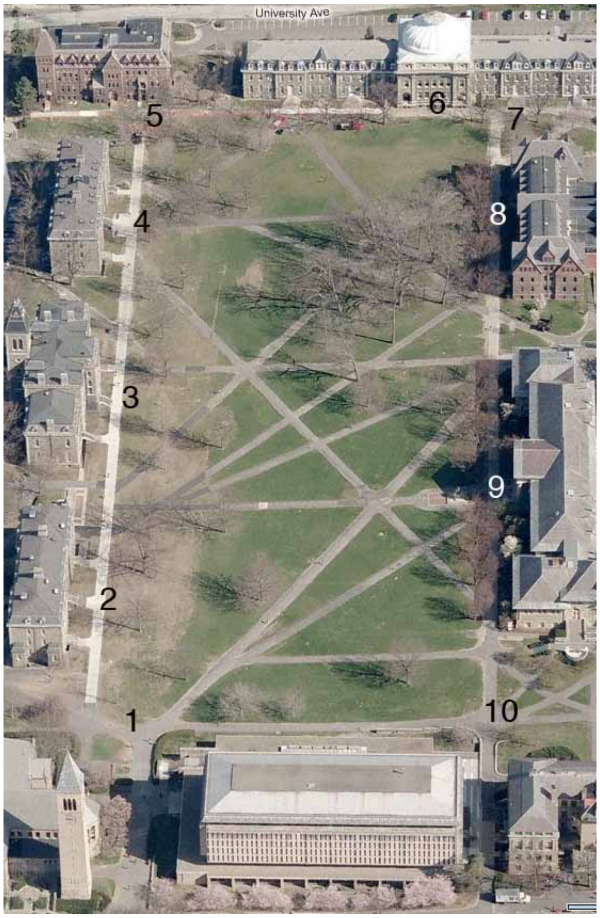

Figure 1: Cornell Arts Quad

Let the set of nodes be $A = \{x | x \in \{1, 2, \ldots, 10\}\}$. Now **we can define a path to be an ordered pair** (a, b) and the set of all paths as the relation

$$R = \{(a, b) | a, b \in A, a \neq b\} \tag{1}$$

since every pair of distinct nodes will define a line segment, or one-way path, in the plane. Then the set of all possible configurations of paths is given by the power set $P(R)$. This set has 2^{90} elements (the cardinality of a power set of a set with 90 elements)

This presents an overwhelming set of possibilities, but fortunately there are three constraints, which we imposed to make our set less unwieldy. We will only model:

1. Non-directed graph: Currently the space $(1, 3)$ is distinct from the path $(3, 1)$. We find this to be unreasonable as pedestrian paths are very rarely "one-way"

2. Connected graph: Aesthetically and functionally it makes little sense to allow a building to be surrounded completely by grass. Furthermore, we picked the ten nodes because we considered them to be essential circulation points of the quad. Therefore, having one of them disconnected from the network would be unreasonable

3. Graphs including the perimeter: This is again chosen in line with our opinions on aesthetics and utility. While pedestrians are likely to accept longer distances than a straight line to remain on the official path, it seems unlikely that a person going from A to B will abide with a path that strictly increases the distance to B before allowing the pedestrian to actually approach B (see figure 2 below).

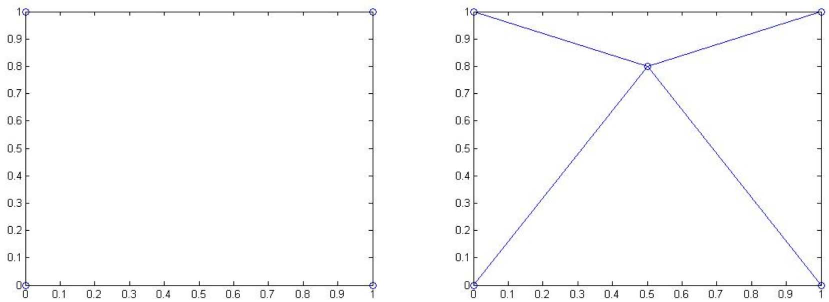

Figure 2: In the first graph, we see that travelling from $(0,1) \to (1,0)$ never increases the distance from $(1,0)$, whereas in the second graph, going from $(0,0) \to (1,0)$ incurs this cost

Once we have attained potentially optimal configurations that satisfy these three constraints, we will remove the constraints and show that there is little or no improvement to be gained by relaxing them. Notice that the third condition implies the second in the geometry that we have chosen, where all nodes are placed on the perimeter. However, in a geometry where nodes in the interior are considered as well, this implication does not hold. Therefore, we have made both constraints explicit.

If we take $B \subseteq P(R)$ to be the set of all possible configurations of paths, which do not satisfy $(1)-(3)$, our model considers the set $P(R)\backslash B$. Having thus define the set of possible configurations of paths we can uses our site-specific geometric information to describe the important features of a given configuration of paths.

Part III

The Length of a Path

A given configuration of paths ($a \in P(R)\backslash B$) and the location of its nodes (see Table 1 below) defines a set of line segments representing paths in the plane.

Node	1	2	3	4	5	6	7	8	9	10
Location	$(0,0)$	$(0,3)$	$(0,9)$	$(0,15)$	$(0,18)$	$(5.6,18)$	$(7,18)$	$(7,14)$	$(7,6)$	$(7,0)$
	SW	Morrill Hall	McGraw Hall	White Hall	Tjaden Hall	Sibley Hall	NE	Lincoln Hall	GS Hall	SE

Table 1: Locations of the Nodes in the Cartesian Plane

Employing simple geometry we can determine l_i, the length of the i^{th} path as a function of the i^{th} path p_i which is given by the coordinates of the two nodes it connects. Therefore we have:

$$l_i = \sqrt{\sum_i p_i^2} = ||\vec{p}||_2 \tag{2}$$

$$a = \{p_1, p_2, \ldots, p_n\}\ p_i = (a,b) \tag{3}$$

where a, b are the numbers of nodes with coordinates (a_1, a_2), (b_1, b_2) respectively. The total length of the paths is then given by summation over i:

$$L = \sum_i l_i \tag{4}$$

By definition, we can define a metric on the graph by:

$$||L|| = \inf_i L = \inf_i \sum_i l_i \tag{5}$$

where L is the minimal length of all the paths.

Part IV

Unofficial Path Induced by Human Behavior

Functionality is a key aspect of any design. To measure the functionality of a path configuration, we have to be able to model the behavior of the people who will use it. A failure to be functional results in both a waste of asphalt and a traffic-damaged lawn.

The problem statement provides us with the following information to model human behavior:

- If the path between two points is 15% longer than the straight line connecting the points, a pedestrian will leave the path and cut across the quad
- An average pedestrian might leave the path if it implies saving more than 10% of the total length of the path

So if a given configuration of paths is too sparse[2], people will choose to save time and cut across the lawn. Notice that the human behavior we are modeling considers **the quad as a space that is to be traversed as quickly as possible**. In other words, we are ignoring traffic from people who go to points on the quad, such as sunbathers or frisbee players. Therefore we can generalize our assumptions of human behavior to be the following:

- Individuals view time spent walking as a detriment to their utility. They try to cross the quad as quickly as possible
- Individuals do not wish to deviate too greatly from sociological norms. Therefore, we are able to assume that there will be a certain propensity to use an official path even though the distance it entails is longer than a straight line path to the destination.

The second condition is discussed in models of human trail formation[3] in which individuals decide how to choose a path based on the collective behavior. Students want to get where they are going, yet they may temper this desire if it makes them the lone-man in the middle of the field.

So an **individual's behavior will be a function of the path configuration**. We interpret "might," as stated in the information given in the problem statement, as a 50% chance that an individual will leave the map finally, if a path coincides with the straight line between two nodes, no utility maximizer would have incentive to leave the path. We fit these three data points with a logistic curve which then describes the propensity of an individual to cut as a function of the relative cost incurred by staying on an official path. The amount of damage done to the grass will then be given by multiplying this propensity to cut by the length of the straight line path.

We determine z, the **relative incurred cost of remaining on an official path** by comparing Dijkstra's algorithm[4] on two graphs: any configuration of paths being considered and the path

[2] A sparse graph is a graph in which every subgraph has a number of edges that is far less than the maximal number of edges in the graph, [2]

[3] See [5],[6],[7]

[4] An algorithm that solves the single-source shortest path problem (see [2],[3]). The variant that we used in particular is [9]

configuration in which all nodes are connected to all other nodes, the complete graph. Once z is determined for every pair of nodes, the logistic function g gives the propensity of an individual to leave the official path for each of these pairs. The length of cut paths can again be obtained via Dijkstra's algorithm. Finally the total damage, C_{subcut} by path-deserting traffic can be obtained by taking the product of the propensity to cut, $g(z)$, with the length of the cut path, summed over all paths.

Part V
The Cost Function

We proceed by identifying the two main costs of a configuration of paths:

1. Maintenance costs for path
2. Landscaping costs for treating unofficial paths

Thus the overall cost of a design becomes a function of both form and functionality. Our first cost penalizes simply paving the entire quad, which is suggested by the problem statement. The second prevents us from leaving it as an unadulterated pasture. For simplicity we define the cost function C, to be a simple linear combination of these two costs:

$$C(L, C_{subcut}) = L + kC_{subcut} \tag{6}$$

where k is the relative per length cost of maintaining asphalt to reviving trafficked grass.

Thus given a configuration of paths on the Arts quadrangle, we can evaluate its cost. We define an **optimal configuration** as one that will minimize the cost in addition to satisfying the following constraints:

1. The configuration should not call for paths which intersect a tree or statue
2. The configuration will not result in lawns with sharp angles or wild geometries
3. The configuration will not make the quad overly fragmented, as we want to prevent a high number of contiguous lawns

Optimally, these additional constraints would be imposed on the domain of our cost function $P(A)\backslash B$ and thus reduce the space over which we must search for an optimum solution. However, we believe that these constraints will not substantially reduce the set of possible paths because:

- We are considering a finite number of possible paths and are not attempting to span the quad with randomly-generated paths. If the path between two of our fixed nodes intesects with a tree, we can simply move one of the nodes
- The complete path configuration does not result in an overwhelming geometry or extremely sharp angles
- The complete path configuration does not fragment the quad to an unreasonable amount

Thus in order to save time required for computing the new domain, we impose that the optimal solution we find must satisfy the above constaints, post-processing. If it does not, we select the next best solution in our set of optimum solutions[5] that satisfies the constraints.

[5]Since we are randomizing the paths we are taking, we can always make the solution set equal to the optimal solution $\pm k\sigma$, where σ is the standard deviation of the probability distribution we impose on our Markov chains. See [3]

It is possible that one may want to include these citeria as part of the cost function. While it is possible to determine the number of contiguous regions determined and the angles of those regions, it is difficult to say what is "too sharp" or "too fragmented." Nonetheless, our method is largely compatible with modifying the cost function if an objective measure of these constraints was known.

Though we have significantly reduced our domain of consideration, it is still quite large. In addition our cost function is expensive to evaluate as it is a nonlinear function of 90 variables (10 nodes $\cdot$ 10 nodes = 100 path - 10 (paths to self) = 90). As a result we cannot hope to perform an optimization analytically or via methodically exploring all of $P(R)\backslash B$. We therefore turn to two alternative methods outline in the section $VIII$.

Part VI

Assumptions:

Assumption	Implication	Justification
The ten nodes on the perimeter offer a reasonable description of the general traversals of the quad	We confine our search for the optimum to path configurations, which connect these ten points.	We chose the points strategically in front of the entries of building positions and at the corners of the quad.
A person going from A to B will not abide with a path that strictly increases the distance to B before allowing the pedestrian to actually approach B	All configurations include the perimeter	Intuition
People trying to traverse the quad form one edge to another	We do not consider nodes in the interior of the quad	The majority of people cross the quad
The word "might" as stated in the problem statement is assumed to mean a 50% chance that an individual will leave the path	Define quantitatively using the Logistic function	Organizations such as the IPCC use similar interpretations
The cost function, C, is a simple linear combination of the path maintenance cost and the landscape maintenance cost	Our definition of the cost function.	Simplicity

Part VII

Algorithms for Finding Optimal Path Configurations

We attempt to find our optimal solution via two methods:

- Brute Force Sampling - generated $30,000$ random configurations from which we select the lowest cost solution as the optimum. The random configurations were generated from the set of all possible pairs of nodes (i.e. R) using the $rand()$ function in Matlab[6]
- A Primitive Version of a Markov Chain - Monte Carlo (MCMC) method

[6] We acknowledge the inaccuracies in the random number generator in Matlab; however, due to time restrictions we were forced to use the built-in random number generator. Luckily, swapping out the Matlab random number generator for another random number generator that is coded in Matlab is a simple exercise

- Inspired by much more careful and effective optimization methods, we developed a primitive MCMC method which mimics several key features of these algorithms. We begin by randomly selecting an initial configuration for the quad's walkways. From this current state we randomly perturb (remove or add) a random number of edges (between 1 and 5). The cost of this new state is then evaluated and compared to the current state. If the new state is superior to the current state, it replaces the current state. If it is an inferior state, it will replace the current state with a 15% probability and is discarded otherwise. Again, as in the brute force sampling, we restrict our search domain to $P(R)\backslash B$.

Part VIII

Results:

We find ,as expected, that our primitive MCMC search yields a better minimal cost solution than those found in the brute force approach. Along with the two solutions obtained via the MCMC approach we consider the 10 best solutions from the brute force search. Using a similar procedure as in the MCMC algorithm, we randomly perturbed each of these solutions by one edge multiple times. If in any single perturbtion we find that the cost improves, the new state is recorded.

As expected, the solutions that were found as a result of the brute force algorithm were not all stable under perturbation. Four of these 10 solutions were improved and of these none had a cost reduction greater than 15%. Encouragingly, neither of the two results obtained from the primitive MCMC search were improved in 1000 random perturbations. This indicates that these solutions may be a local minimum, if not a global minimum of the cost function in the domain we consider (See Figure 3 below). From these two solutions we chose one to recommend based on which solution best satisfied the subjective constraints discussed on part II. In the absence of an objective way to evaluate a solution based on these criteria as part of the cost function we believe that this method of democratic voting by the concerned parties is the most reasonable alternative.

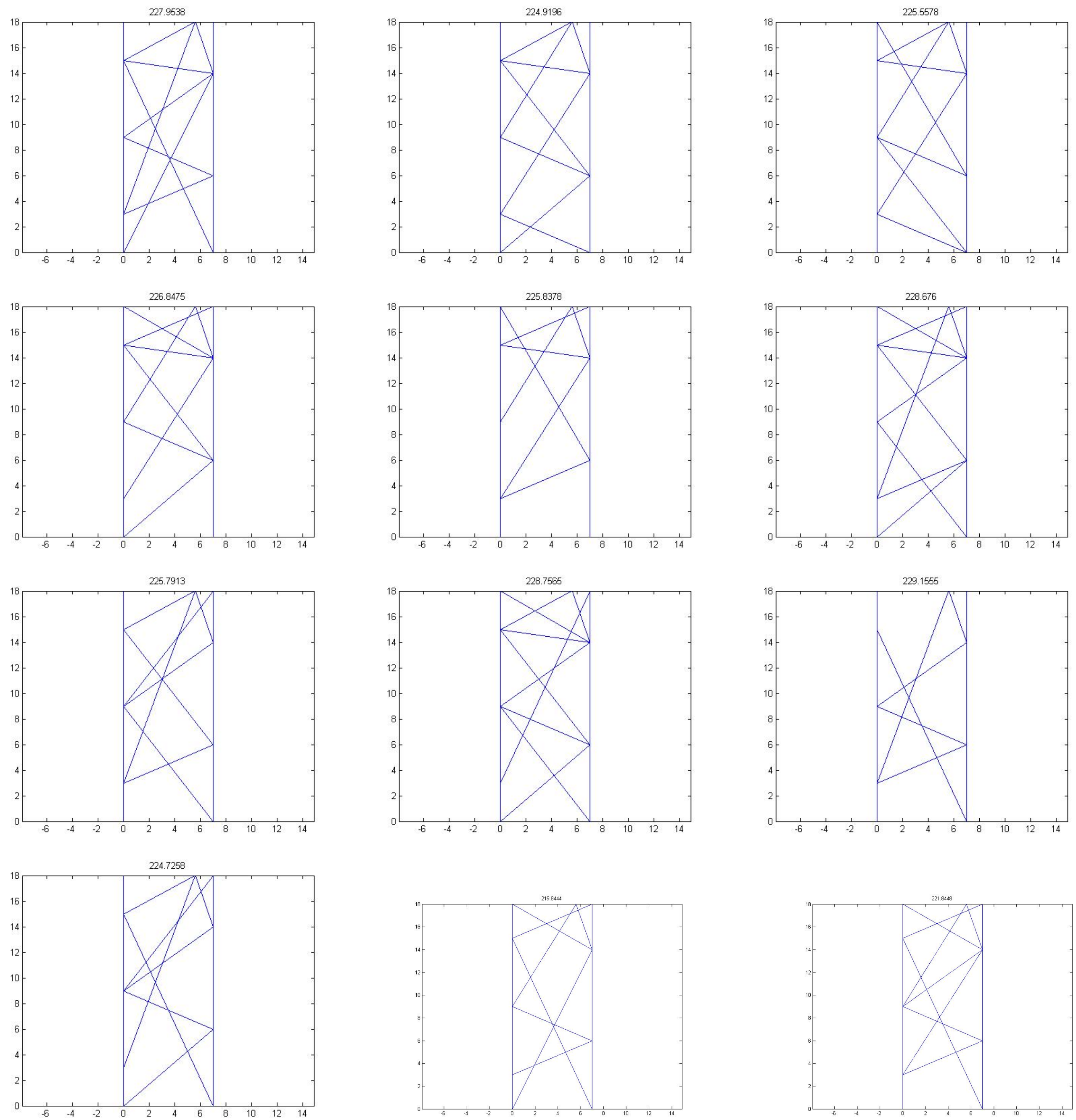

Figure 3: 12 optimal solutions; the first 10 depict the solutions for the brute force method, whereas the last two depict solutions for the MCMC method. The cost is the title of each graph. Note: We chose the second to last graph as our optimal solution

Part IX

Recommended Solution

The network of paths given below would most likely minimize the costs associated with path and lawn maintenance:

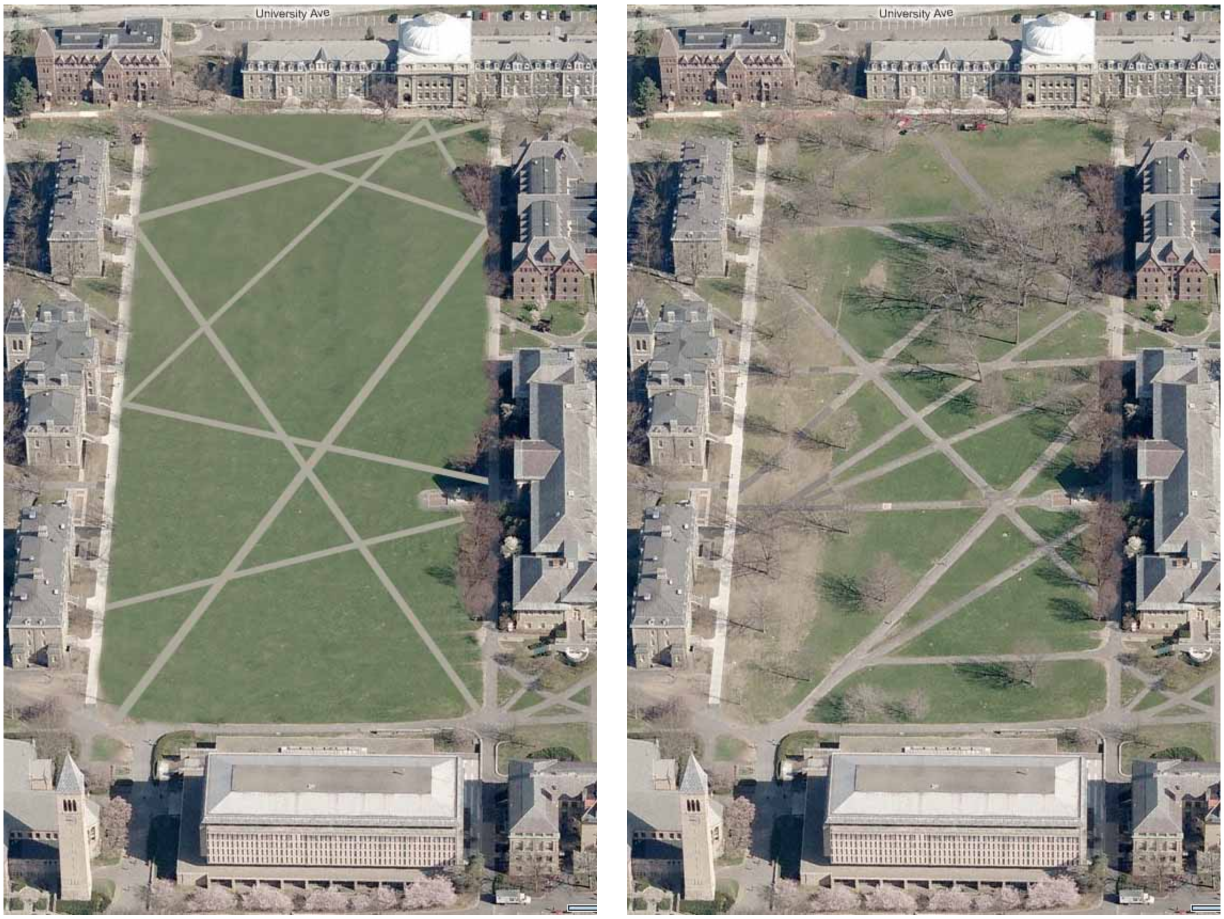

Figure 4: Our solution, left, and the current Arts quad

We also find, in line with our predictions, that the solution does not intersect any trees of statues. This is depicted below in figure 5:

Figure 5: Our plan in red, which does not intersect with any trees or statues

Using our model we can then determine to what extent these paths will lead to unofficial trail formation. We find that these formations are minimal, and this is expressed in the following figure which displays the new quad after significant use.

Figure 6: Our optimal solution, with the (estimated) unofficial, treaded paths drawn in. Their opacity is proportional to the propensity to cut, as defined in section 2

Our final reccomendation is for the placement of special hardier seeds. We reccomend that these seeds be used in the highlighted regions where unofficial pathways are most detrimental and where large continuous areas will attract frisbee players and the like. This is portrayed below:

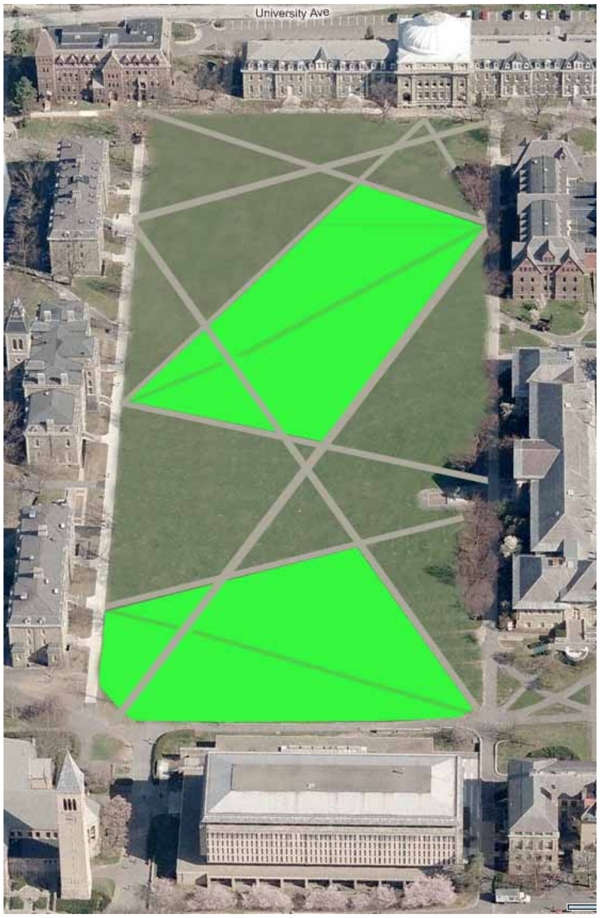

Figure 7: Our recommendation for where the more expensive seeding should be placed

Part X
Future Work

The flexibility of our approach gives rise to a copious number of future directions in which our method could be extended. Here we catalog several possibilities:

1. Develop or find a proper optimization routine: Clearly before proceeding to complicate the above scheme, it would be best to have an appropriate method to perform the optimization crucial to our analysis.

2. Add additional nodes: In some sense we have drastically limited the space of possible solutions by demanding that all nodes of our graph live on the perimeter of the graph. In addition to simply considering adding a large set of internal and perimeter nodes, we note that certain special additions, such as Steiner points, may result in large cost reduction.

3. In this analysis we have divided the important features of the system into two categories, conditions and constraints. Often for the sake of avoiding subjective comparisions, we have described features such as aesthetic appeal or angularity of the quad as constraints. While we believe objectively determining the relative importance of such characteristics would be difficult, there is **no prohibition in expanding the considerations of our cost function**. Once established the

optimization with respect to a cost function could proceed as before.

4. The current description of human behavior contains perhaps the most simplifications made by our model. There already exist several models[7] which quantitatively reflect the collective behavior of humans with respect to trail fomation. Employing these models would result in a better estimate of the maintenance costs.

5. We have assumed that the perimeter of the quad should necessarily have paths. An important test of this assumption is to relax it and perturb any optimal solutions to see if the relaxation will result in a more cost effective solution.

6. Finally, we have assumed that our optimal solution will have a straightline geometry. The use of curved walkways may provide significantly less expensive configurations.

Part XI

Bibliography

[1] Strogatz, S.H.. "Exploring Complex Networks." Nature 438(2001): 268-276.

[2] Roberts, John, S.. Graph Theory and Its Applications to Problems of Society. SIAM, 1987.

[3] Gamerman, Dani. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. CRC Press, 1997.

[4] Gross, Jonathan, and Yellen, Jay. Graph Theory and Its Applications. Boca Raton: CRC Press, 1999.

[5] Helbing, Dirk, Schweitzer, Frank, Keltsch, Joachim, and Molnar, Peter. "Active Walker Model for the Formation of Human and Animal Trail Systems." Physical Review. 56(1997): 2527-2539.

[6] Helbing, Dirk, Keltsch, Joachim, and Molnar, Peter. "Modelling the Evolution of Human Trail Systems." Nature 388(1997): 47-50.

[7] Helbing, D. Quantitative Sociodynamics. Stochastic Methods and Social Interaction Processes (Kluwer Academic, Dordrecht, 1995).

[8] Dynkin, Eugene. Markov Process: Volume I, 1st ed. Berlin: Springer-Verlag, 1965.

[9] Kirk, Joesph. "Djikstra's Algorithm." May 2008, Matlab Central.

[10] Holz, Sebastian. "Curve Intersection." November 2005, Matlab Central.

[7] See [5],[6],[7]