Facebook friend wheels and quadratic assignment

Cody Bredendick   Michael C. Ferris

University of Wisconsin, Madison

INFORMS National Meeting, San Diego: October 11, 2009

Teaching optimization and modeling

- Optimization Tools and Modeling Course: senior level undergraduate
- Lots of material, lots of modeling exercises
- Required final project: individual, student defined
  - Highlights difficulties of data acquisition
  - 1 page proposal, oral interview
  - 4 page summary report (plus additional models that may or may not be exercised)
  - Always impressed by a significant fraction of these student driven idea projects - recommend process to all

Conformal Radiotherapy
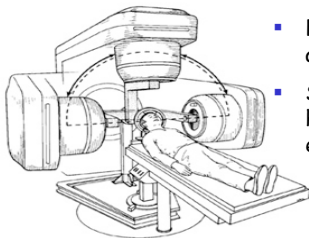
- Fire from multiple angles
- Superposition allows high dose in target, low elsewhere
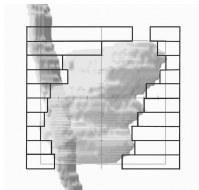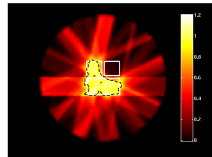
- Beam shaping via collimator
- Gradient across beam via wedges

Biological Pathway Models

Opt knock (a bilevel program)

max   bioengineering objective (through gene knockouts)

s.t.   max cellular objective (over fluxes)

       s.t.   fixed substrate uptake

             network stoichiometry

             blocked reactions (from outer problem)

       number of knockouts $\leq$ limit

Standard model: The goal of this project is to build a model that accounts for the "gene" deletions and regulatory interactions, and solve the resulting problem for different biochemical productions to identify diverse metabolic engineering strategies. The initial model includes about 30,000 constraints and 20,000 variables (10,000 binary).

Video stabilization

- Videos shot with a hand-held conventional video camera often appear remarkably shaky
- Camera shake is one of the biggest components of the large quality difference between home movies and professional videos
- Several commercial software packages are available for digitally removing or reducing camera shake as a post processing step
- Opaque, open source version of this

Nonlinear least squares:

$$\min_{\{H_f\}_{f=1}^F} \sum_{f=2}^{F-1} \sum_{j \in \phi_f} \left\| u_{f,j} - \frac{1}{2}(u_{f-1,j_{prev}} + u_{f+1,j_{next}}) \right\|^2 + \Phi(\{H_f\}_{f=1}^F)$$

where

$$\Phi(\{H_f\}_{f=1}^F) = \sum_{f=2}^{F-1} \lambda_1 \theta_f^2 + \lambda_2 t_{x,f}^2 + + \lambda_3 t_{y,f}^2 + \lambda_4 s_{x,f}^2 + \lambda_5 s_{y,f}^2$$

H is the unknown transform, parameterized by θ, t and s.

Airline network bid prices

- Talluri and Van Ryzin (randomized LP) model to compute network bid prices
- Data from hub and spoke airline (LAN airlines)
- 220 flight legs, 14330 itineraries (using iterated SQL self-joint queries), 20 day booking horizon
- Demand estimation (randomly generated)
- Results are less favorable than those of Talluri and Van Ryzin - student gives several possible places where errors could exist...

Facebook

- No introduction needed: one of the largest and most visited websites around
- Practically defined the modern definition of "social networking"
- New features constantly developed
- Facebook released a developers' API: allows apps to be developed by anyone
- A friend wheel is one such app

A friend wheel
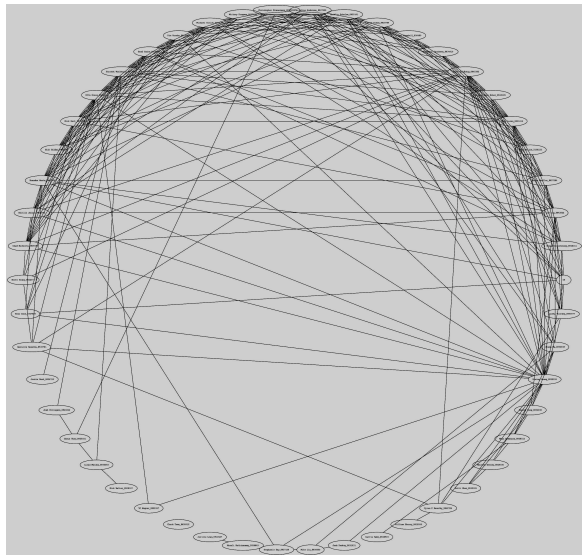
A friend wheel

- An image used to provide a visual representation of the relationships between the friends any one person may have
- constructed by gathering each of the target individual's friends and placing them equidistant from each other on the circumference of a circle
- line segments are drawn between each point (each friend) if the people that those two connecting points represent are friends with each other
- Ordering determines resulting image
- With random ordering, one can expect a very dense mesh of lines equally distributed across the area of the circle
- Inspiration of project
- Order to reduce amount of ink used

A friend wheel
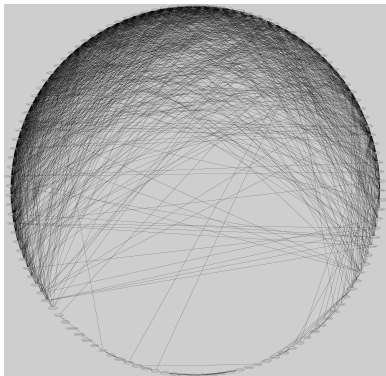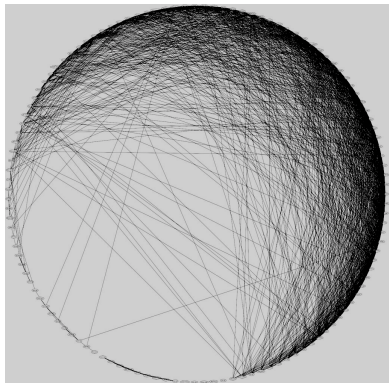
Typical size from 50 to 1000 friends

Overall scheme

- Gather the data
  - Build a small web service to retrieve data
  - Manipulate data to input format for optimization solver
  - Implementation: via php scripts within Facebook
- Determine the ordering
- Output the image
  - Manipulate optimization output for image package
  - Use graphviz package (publically available)
  - Two passes, one to fix locations on circle and draw connecting arcs, second to generate image in nice format (png, jpg, pdf)

QAP (Koopmans and Beckman)

Given n facilities $\{f_1, \ldots, f_n\}$, n locations $\{l_1, \ldots, l_n\}$:
Determine to which location each facility must be assigned
$W = (w_{i,j}) \in \mathbb{R}_+^{n \times n}$: $w_{i,j}$ represents flow between f_i and f_j
$D = (d_{i,j}) \in \mathbb{R}_+^{n \times n}$: $d_{i,j}$ represents distance between l_i and l_j
$p : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$ is an assignment whose cost is

$$c(p) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} d_{p(i),p(j)}$$

QAP : min $c(p)$ subject to $p \in \Pi_n$

Π_n is the set of all permutations of $\{1, \ldots, n\}$
QAP is known to be strongly NP-hard

Minimizing ink

- n is the number of friends of a given individual
- assign each friend a location on the circle
- $w_{i,j} = 1$ if i is a friend of j, and 0 otherwise
- $d_{r,s}$ is the distance from location r on the circle circumference to location s (Euclidean or otherwise)

Data needed to define this problem is n and the links $w_{i,j}$ that represent that i is a friend of j: compact

Model via GAMS

- define binaries $p(i, r) = 1$ if facility i located at r
- problem is a mixed integer quadratic program

$$\text{min} \qquad \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \sum_{r} \sum_{s} d_{r,s} * p(i, r) * p(j, s)$$

$$\text{subject to} \qquad \sum_{i} p(i, r) = 1, \forall r$$

$$\sum_{r} p(i, r) = 1, \forall i$$

- codes go nowhere

Problem is not necessarily convex. Fix one location to remove symmetry.

Model via GAMS

- additionally define binaries $q(i,r,j,s) = (p(i,r)$ and $p(j,s))$
- problem is a mixed integer linear program

$$q(i,r,j,s) \leq p(i,r)$$
$$q(i,r,j,s) \leq p(j,s)$$
$$q(i,r,j,s) \geq p(i,r) + p(j,s) - 1$$

- codes go nowhere

Well known that exact solution is very hard, even for small ($n = 40$) problems

Heuristic solution

- Many heuristic solution methods for such problems (see Voss survey)
- Greedy randomized adaptive search procedures (GRASP) are one type (Pardalos, Resende, et al)
- Source code available for download

- generate starting solution using greedy heuristic
- apply local search until termination criteria satisfied
- apply "path-relinking" (heuristic to combine to good solutions)

Some results

solver	options	25		50		125	
		time	soln	time	soln	time	soln
dicopt		1.3	195	27.0	1752		
sbb		4.6	189	98.9	1742		
qsolver		0.2	186	1.3	1660	16.7	25107
	-npr					7.0	25143
	-npr -n 2					14.0	25143
	-npr -i 64					14.0	25143
	-npr -i 10					2.2	25143
	-npr -i 5					1.1	25398
	-npr -s 250					7.1	25144

Details of procedure

- Login to your facebook account
- Navigate to http://apps.facebook.com/cscbfriend/
- This page calls index.php on a remote "execution" websever
- Initiates a non-blocking call to start the data gathering process
- php preparedata.php [userID] [sessionID]
  - ▸ Communicates with facebook API
  - ▸ Generate a list of friends, and then a list of connections (using the Facebook Query Language)
  - ▸ Loops over data in order to generate names.txt, arcs.txt, qinput.txt
  - ▸ Executes a shell script to run "qsolver" and "graphviz"

Timer program (hack) to allow page more than default time to load.

The competition
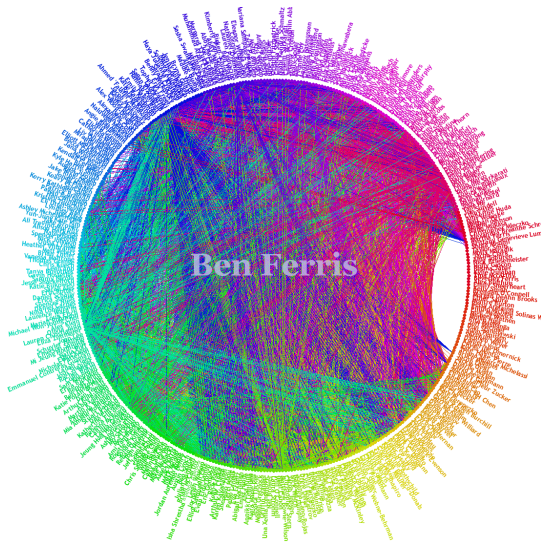
Output is slow to generate, not as good as QAP solution, but resulting graph is not static - colored, much more sexy

Aim to collaborate to switch out the solver engine to use sophisticated optimization tool under the hood

Conclusions

- Try it at: http://apps.facebook.com/cscbfriend/
- Optimization critical in application, but development time dominated by data handling
- Interesting source of new test programs for QAP - solution time still dominated by optimization
- Modern optimizers must have a range of skills - modeling in multiple formats, ability to merge model solutions, skills to extract data and display results