The perfect conference
Using stochastic optimization to bring people together

Brian Lange @ Strata San José 2017
who is this guy?
who are these folks?
Today

- Case study: our work with Research Corporation for Scientific Advancement (RCSA)

- Some high level explanation of the technique we used (simulated annealing)

- Our vision of the role of a data scientist, and how they interact with the people they’re serving
Meet Richard*

*(bad) artist’s rendering
Richard’s needs

“Some of the goals of Scialog conferences are to facilitate the formation of new collaborative teams, encourage sharing insights and catalyze novel lines of research.”
What’s different about Scialog

- Cross-disciplinary topics
- Small conferences (40–50 junior scientists, 10–15 senior scientists)
- Invite only
- Unconventional format

How well does this work?
Quantifying connection
Quantifying connection
Quantifying connection

unaware → aware → discussed
Quantifying connection

unaware → aware → discussed → collaborate
Quantifying connection

<table>
<thead>
<tr>
<th>Name</th>
<th>Unfamiliar</th>
<th>Awareness</th>
<th>Discussion</th>
<th>Collaborator</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Names redacted]</td>
<td>[Check marks]</td>
<td>[Check marks]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Check marks]</td>
<td>[Check marks]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Blank]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
<tr>
<td>[Names redacted]</td>
<td>[Blank]</td>
<td>[Check marks]</td>
<td>[Blank]</td>
<td>[Blank]</td>
</tr>
</tbody>
</table>
before

after
Scialog 2015: Molecules Come to Life

- 118 new conversations
- 17 new collaborations
- 100% growth from pre-conference
- 53.6% of attendees formed a new collaboration with an attendee
measure $\rightarrow$ improve
Mini Breakout Groups - round 1

conference attendees

A  D  G  J
B  E  H  K
C  F  I  L
Mini Breakout Groups - round 2

- Group 1: B, D, G, L
- Group 2: C, H, I, K
- Group 3: A, F, J, E
Mini Breakout Groups - round 3

- Group 1: J, D, F, L
- Group 2: E, I, G, K
- Group 3: A, H, B, C
Are these good groups?

- Group 1: J, D, F, L
- Group 2: E, I, G, K
- Group 3: A, H, B, C
Is this a good group?
Is this a good group?

terribleness: 7
Is this a good group?

terribleness: 2
possible configurations

$1.47 \times 10^{40}$, or 14 duodecillion
quadrillions of years :(
optimization

- terribleness
- slope
- intercept

Diagram showing optimization process:
- Learning step
- Minimum
- Random initial value
terribleness: 7
terribleness: 2
terribleness: 6
terribleness:1
terribleness: 8
enter:
simulated annealing
simulated annealing

* a type of stochastic optimization

stochastic optimization

algorithm with randomness introduced to avoid getting caught in local minima
1. make a random “move” from current state

2. check if the new state is better than the previous

3. If it is, keep it and skip to 5.

4. If not, there’s a chance we’ll still keep it, depending on the temperature

temperature: 100
terribleness: 350
previous terribleness: 350
1. make a random “move” from current state

2. check if the new state is better than the previous

3. If it is, keep it and skip ahead.

4. If not, there’s a chance we’ll still keep it, depending on the temperature
1. make a random “move” from current state

2. check if the new state is better than the previous

3. If it is, keep it and skip ahead.

4. If not, there’s a chance we’ll still keep it, depending on the temperature

5. Stop when temperature is too low. Otherwise, reduce temperature and repeat.
1. make a random “move” from current state
2. check if the new state is better than the previous
3. If it is, keep it and skip ahead.
4. If not, there’s a chance we’ll still keep it, depending on the temperature
5. Stop when temperature is too low. Otherwise, reduce temperature and repeat.

temperature: 90
terribleness: 243
previous terribleness: 380
1. make a random “move” from current state

2. check if the new state is better than the previous

3. If it is, keep it and skip ahead.

4. If not, there’s a chance we’ll still keep it, depending on the temperature

5. Stop when temperature is too low. Otherwise, reduce temperature and repeat.
fast forward
1. make a random “move” from current state

2. check if the new state is better than the previous

3. If it is, keep it and skip ahead.

4. If not, there’s a chance we’ll still keep it, depending on the temperature

5. Stop when temperature is too low. Otherwise, reduce temperature and repeat.
1. make a random “move” from current state
2. check if the new state is better than the previous
3. If it is, keep it and skip ahead.
4. If not, there’s a chance we’ll still keep it, depending on the temperature
5. Stop when temperature is too low. Otherwise, reduce temperature and repeat.

temperature: 3
terribleness: 194
previous terribleness: 63
1. make a random “move” from current state

2. check if the new state is better than the previous

3. If it is, keep it and skip ahead.

4. If not, there’s a chance we’ll still keep it, depending on the temperature

5. Stop when temperature is too low. Otherwise, reduce temperature and repeat.
Not just for groups of things

- circuit board designs
- stock trading rules
- **anything** which has a defined state, a “move” (aka transition to another state), and a measure of goodness
very, very tweakable

- make “smarter-than-random” moves
- change the cooling function (reheat and cool)
- save best state so far and restart from it periodically
- repeat many times from different random starting states
- etc...
very, very tweakable

- make “smarter-than-random” moves
- change the cooling function (reheat and cool)
- save best state so far and restart from it periodically
- repeat many times from different random starting states
- etc…
let’s talk more about terribleness
client needs

- minimize how “connected” the people in each group are
- don’t want people to be in the same group with somebody twice
- etc...
iterative problem solving
anticipate refining your solution

generate ideas
“what does Richard need?”

evaluation
build prototype
iterative problem solving

anticipate refining your solution

generate ideas

evaluation

build prototype

“these could be better”
iterative problem solving

anticipate refining your solution

generate ideas

"let's tweak the optimization technique"

evaluation

build prototype
iterative problem solving

anticipate refining your solution

“much better, what do you think Richard?”
iterative problem solving

anticipate refining your solution

“pretty good, but I don’t like ______”
iterative problem solving

anticipate refining your solution

generate ideas

“cool, let’s tweak the objective function”

evaluation

build prototype
iterative problem solving

anticipate refining your solution

generate ideas

evaluation

build prototype

“these look great!”
iterative problem solving
also an optimization algorithm of sorts...

“these are great!”
so, did it work?
Scialog 2015: Molecules Come to Life

- In **30 out of 39** group discussions (77%), **none** of the people in those groups had even heard of each other before

- In **all** group discussions, **no one had spoken or collaborated with anyone else in their group**

- On average, a group had **2.6** different disciplines (physics, biology, etc)

- Each group had **at least one theorist and one experimentalist**
Scialog 2015: Molecules Come to Life

_______ is important to the success of Scialog

![Bar chart showing Mini Breakouts and Regular Discussions]

- Mini Breakouts
- Regular Discussions

Very Much Disagree | Neutral | Very Much Agree
Scialog 2015: Molecules Come to Life

“We received 20 collaborative proposals, the most ever at a Scialog.”
further iterations
Topic groups/preferences

Tissue mechanics and evolution

J D I F

The physics of genetic networks

K E H A

In vivo mechanics

C B G L
## Topic groups/preferences

<table>
<thead>
<tr>
<th></th>
<th>No way</th>
<th>Might nap</th>
<th>Would listen</th>
<th>Would chime in</th>
<th>Really into it</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Topic B</strong></td>
<td></td>
<td></td>
<td></td>
<td><img src="https://via.placeholder.com/35" alt="Circle" /></td>
<td></td>
</tr>
<tr>
<td><strong>Topic C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="https://via.placeholder.com/35" alt="Circle" /></td>
</tr>
</tbody>
</table>
## Desired connections

<table>
<thead>
<tr>
<th>Name</th>
<th>I'd like to have detailed discussions with this person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value 1</td>
<td>☑</td>
</tr>
<tr>
<td>Value 2</td>
<td></td>
</tr>
<tr>
<td>Value 3</td>
<td></td>
</tr>
<tr>
<td>Value 4</td>
<td></td>
</tr>
<tr>
<td>Value 5</td>
<td></td>
</tr>
<tr>
<td>Value 6</td>
<td></td>
</tr>
<tr>
<td>Value 7</td>
<td></td>
</tr>
<tr>
<td>Value 8</td>
<td></td>
</tr>
<tr>
<td>Value 9</td>
<td></td>
</tr>
<tr>
<td>Value 10</td>
<td></td>
</tr>
</tbody>
</table>
### Nominated topic experts

**Nominating critical discussion participants**

Listed below are the topics you expressed interest in. For each topic, if you think another SciLog Fellow is an essential person to have in a discussion on that topic, please indicate them below. You may select up to two for each topic.

<table>
<thead>
<tr>
<th>Topic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Do cells compute in an optimal way? What about organisms? How can we test this?</td>
<td></td>
</tr>
<tr>
<td>Role of cellular heterogeneity: when do differences make a difference?</td>
<td></td>
</tr>
<tr>
<td>How do we measure genetic interactions and how can we use the data?</td>
<td></td>
</tr>
<tr>
<td>Physical basis of stress and disease: how does disease perturb physical properties such as homeostasis and molecular stability?</td>
<td></td>
</tr>
<tr>
<td>Is game theory relevant for evolution?</td>
<td></td>
</tr>
<tr>
<td>Energy production and efficiency in individual cells and organisms</td>
<td></td>
</tr>
<tr>
<td>How can we use landscape ideas in biology? Are they useful?</td>
<td></td>
</tr>
<tr>
<td>What have we learnt so far about the algorithms cells use to make decisions?</td>
<td></td>
</tr>
<tr>
<td>How to measure and model energy production and energy consumption at the cellular level</td>
<td></td>
</tr>
</tbody>
</table>
how might this work at a bigger conference?

- survey including EVERYONE not tenable
- use latent sources like social media
- use RFID or other hardware to infer connections at the conference
- use a survey with a representative sample and similarity metrics to project preferences
how might this work at a bigger conference?

- different format
- separate opt-in discussion track
- rather than groups, make email introductions
what to learn from this
optimization isn’t just for straightforward mathematical problems
data science isn’t just for straightforward mathematical problems
a great data scientist acts as a translator
iteration drives improvement on both methods and data collection
if you can measure it, you have the data to actively improve it
thanks!

questions?