# The Maximum Facility Location Problem
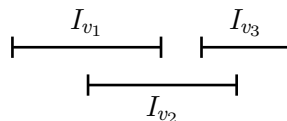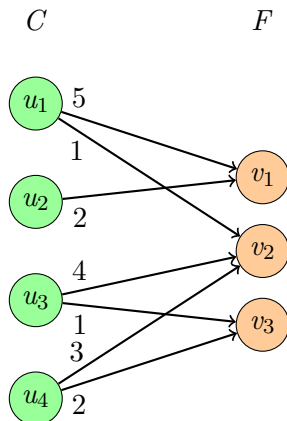
**Mitchell Jones**
**Supervisor: Dr. Julián Mestre**

School of Information Technologies
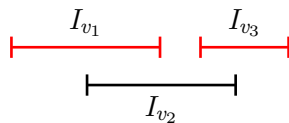Faculty of Engineering & Information Technologies

THE UNIVERSITY OF
SYDNEY

## Example
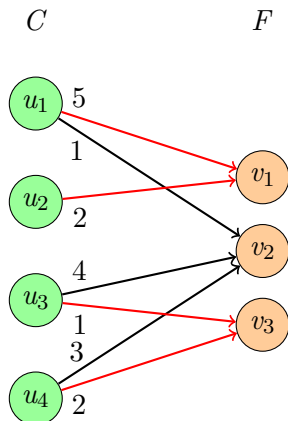
Problem
○●○○

Results
○○○○○○

Application
○○○○○

Conclusion
○○

# Example (cont.)



$C$

$F$

$u_1$ 5

1

$v_1$

$u_2$ 2

$v_2$

4

$u_3$ 1

$v_3$

3

$u_4$ 2

$I_{v_1}$    $I_{v_3}$

$I_{v_2}$

Value $= 10$

## Maximum Facility Location

- $C$ is the set of clients
- $F$ is the set of facilities
- Weight function $w : C \times F \to \mathbb{R}$
- Each facility $v \in F$ has an interval $I_v$ on the real line

### Goal

Select an independent set $S \subseteq F$ maximising cost of assigning clients to facilities.

## Linear Program

$$
\begin{aligned}
\text{maximise} \quad & \sum_{u \in C, v \in F} w_{uv} x_{uv} \\
\text{subject to} \quad & \sum_{v \in F} x_{uv} \ \leq \ 1 \quad \forall u \in C \\
& x_{uv} \ \leq \ y_v \quad \forall u \in C, v \in F \qquad \text{(LP)} \\
& \sum_{v \in F : p \in I_v} y_v \ \leq \ 1 \quad \forall p \in P \\
& x_{uv} \ \geq \ 0 \quad \forall u \in C, v \in F \\
& y_v \ \geq \ 0 \quad \forall v \in F
\end{aligned}
$$

## Overview of Results

- ► Motivated by genome mapping
  - ► Algorithms perform up to $1000\times$ faster
  - ► Comparable precision/recall
- ► Two new algorithms
  - ► Independent rounding: 0.19-approximation
  - ► Dependent rounding: Non-constant approximation
- ► Establish the hardness of the problem

## Independent Rounding: Setup
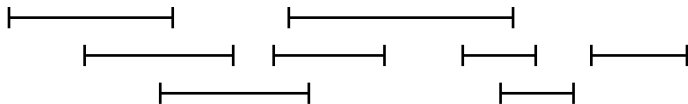
## Independent Rounding: Random Selection

Problem
○○○○

Results
○○○●○○

Application
○○○○○

Conclusion
○○
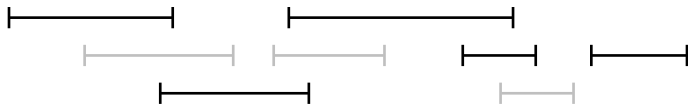
# Independent Rounding: Filtering
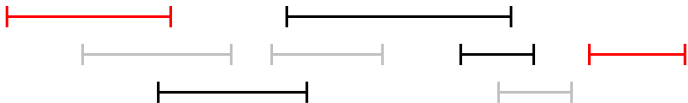
## Independent Rounding: Key Result
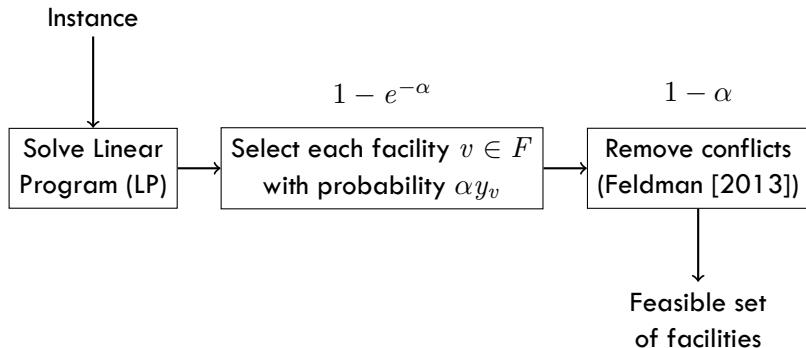


- $(1 - e^{-\alpha})(1 - \alpha)$-approximation
- Choosing $\alpha = 0.44$ leads to 0.19-approximation

## Hardness

### Theorem

For any $\varepsilon > 0$, there is no $(1-1/e+\varepsilon)$-approximation algorithm for maximum facility location, unless P = NP.

## Problem

Problem
0000

Results
000000

Application
0●000

Conclusion
00

## Modelling the Problem
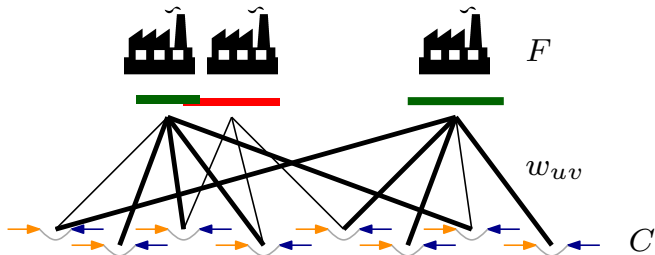
## Experiments
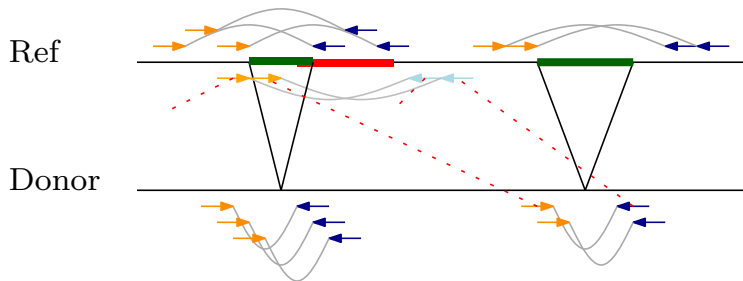
- ► Two data sets
    - ► Craig Venter's genome (simulated)
    - ► NA12878 genome (real data, no ground truth)
- ► Compare precision/recall of predicted vs. actual deletions
- ► Two previous greedy algorithms by Hormozdiari et al. [2010]

## Craig Venter

|             | Recall    | Precision | Runtime (mins) |
|-------------|-----------|-----------|----------------|
| IP          | 65.80     | 40.26     | 10.4           |
| Independent | 65.02     | 40.63     | 15.8           |
| Dependent   | **65.59** | 40.49     | **4.5**        |
| Greedy      | 65.44     | **41.47** | 1840.0         |
| wGreedy     | 65.29     | 41.42     | 4095.0         |

## NA12878

|  | Recall | Precision | Runtime (mins) |
|---|---|---|---|
| 1000G (Mills et al. [2011]) | | | |
| IP | 60.3 | 69.3 | 10.5 |
| Independent | 60.2 | 69.11 | 7.0 |
| Dependent | 60.1 | 69.03 | **3.1** |
| Greedy | 60.2 | 69.79 | 1737.5 |
| wGreedy | **60.3** | **69.99** | 4947.2 |
| long-read (Layer et al. [2014]) | | | |
| IP | 67.1 | 78.9 | 10.5 |
| Independent | **67.4** | 79.5 | 7.0 |
| Dependent | **67.4** | 79.4 | **3.1** |
| Greedy | 67.1 | **80.1** | 1737.5 |
| wGreedy | 66.8 | 79.8 | 4947.2 |

## Concluding Statements

- ▶ Modelled the problem as an integer/linear program
- ▶ Developed two new approximation algorithms
- ▶ Established hardness of problem
- ▶ Biological application is faster than existing methods, without sacrificing solution quality

## Future Work

- An algorithm with better approximation ratio
  - Gap between best known result and ≈0.63
  - Or a tighter hardness bound
- Analyse guarantee of non-constant approximation algorithm
- New conflict resolution heuristics

# Questions?

Thanks to Stefan Canzar & Khaled Elbassioni

## References

Moran Feldman. Maximization Problems with Submodular Objective Functions. PhD thesis, Israel Institute of Technology, 2013.

Fereydoun Hormozdiari, Iman Hajirasouliha, Phuong Dao, Faraz Hach, Deniz Yorukoglu, Can Alkan, Evan E Eichler, and S Cenk Sahinalp. Next-generation variationhunter: combinatorial algorithms for transposon insertion discovery. Bioinformatics, 26(12):i350--i357, 2010.

Ryan M Layer, Colby Chiang, Aaron R Quinlan, and Ira M Hall. Lumpy: a probabilistic framework for structural variant discovery. Genome biology, 15(6):R84, 2014.

Ryan E Mills, Klaudia Walter, Chip Stewart, Robert E Handsaker, Ken Chen, Can Alkan, Alexej Abyzov, Seungtai Chris Yoon, Kai Ye, R Keira Cheetham, et al. Mapping copy number variation by population-scale genome sequencing. Nature, 470 (7332):59--65, 2011.