

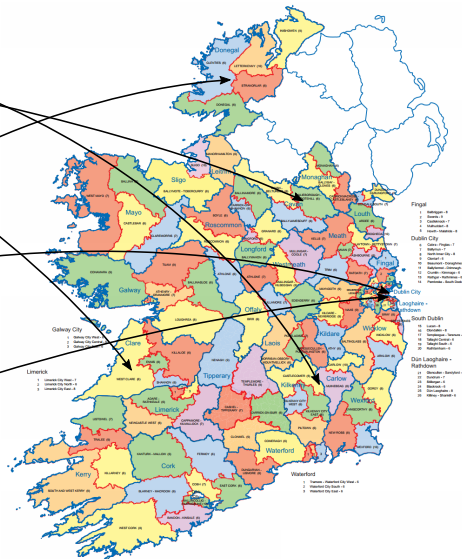
# Complex Clustering Using Constraint Programming: Modelling Electoral Map Creation

Lars Kotthoff, Barry O'Sullivan, S. S. Ravi, Ian Davidson

31 August 2015

## Application: Electoral Map Creation

Constituency	Seats
Carlow-Kilkenny	5
Cavan-Monaghan	5
Clare	4
Cork East	4
Cork North-Central	4
Cork North-West	3
Cork South-Central	5
Cork South-West	3
Donegal North-East	3
Donegal South-West	3
Dublin Central	4
Dublin Mid-West	4
Dublin North	4
Dublin North-Central	3
Dublin North-East	3
Dublin North-West	3
Dublin South	5
Dublin South-Central	5
Dublin South-East	4
Dublin South-West	4
Dublin West	4
Dún Laoghaire	4

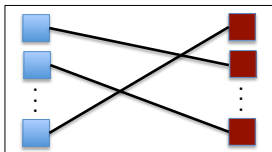


# Motivation for constraint approach

- ▷ hard problem
- ▷ existing approaches use heuristics
- ▷ fast solution, but approximate
- ▷ constraint formulation allows to find global optimum
- ▷ problem only solved every few years, so long runtimes acceptable

# General Pattern: Cluster and Match

Matching Between Clusterings

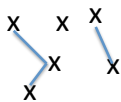


Clustering  
1

Clustering  
m

• • •

Objects



X

Features per object  
Relationships b/w objects  
Constraints



Y

## Problem definition

- ▷ electoral districts to be clustered into constituencies
- ▷ constituencies must be connected
- ▷ representatives to be assigned to (clustered into) constituencies (between 3 and 5 per constituency)
- ▷ make representation as equal as possible, i.e. each representative should represent approximately the same number of people

- ▷ assume representatives are interchangeable – actual clustering does not matter, but only number of representatives
- ▷ cannot compute objective function value without both clusterings – cannot solve clusterings separately and combine
- ▷ electoral districts in one constituency must be connected
- ▷ additional constraints, e.g. no representative may represent more than 30,000 people according to the Irish constitution

- ▷ Boolean matrices for electoral districts and representatives to indicate which cluster they belong to
- ▷ population vector with number of people represented by each representative

## Model: Constraints

- ▷ each electoral district is assigned to exactly one constituency
- ▷ each constituency contains at least one electoral district
- ▷ each representative is assigned to exactly one constituency
- ▷ each constituency has between three and five representatives



## Model: Connectivity

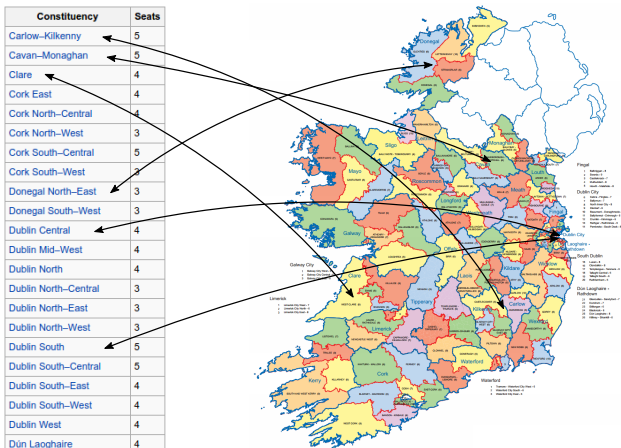
- ▷ cannot have disconnected constituencies
- ▷ for each electoral district, there must exist a path to every electoral district in the same constituency that traverses only districts in that constituency
- ▷ encode as tree constraint – the districts in every constituency form a tree
- ▷ connectivity in tree is based on adjacency of districts
- ▷ root node connected to itself
- ▷ exactly one root node per constituency

## Model: Objective

- ▷ population a representative represents is the total population of the constituency the representative is assigned to divided by the number of representatives assigned to this constituency
- ▷ minimise difference between minimum and maximum population per representative

- ▷ large solution space (3409 electoral districts, 166 representatives, 43 constituencies)
- ▷ weak constraints
- ▷ encoding of connectivity large
- ▷ ground truth (current situation) not completely available

- ▷ implementation in Numberjack
- ▷ best results with MIP translation and Gurobi solver
- ▷ very challenging problem – takes extremely long to solve
- ▷ currently unable to solve entire problem
- ▷ even subsets of the map are very challenging



- ▷ model electoral map creation as constraint problem to find global optimum
- ▷ scaling challenges when solving
- ▷ ideas for making the model better?