Sustainable Places Research Institute
Department of Geography and Planning

Some applications for Spatial Network Analysis: transport, accessibility, health, communities, economics and classification of regions

Crispin Cooper

Feat. Scott Orford, David Fone, Alain Chiaradia, Ian Harvey, Aamir Mohammed, Anwar Hossein, Richard Price, Chinmoy Sarkar, Brian Webb, Neil Harris, Adrian Healey

Office for National Statistics

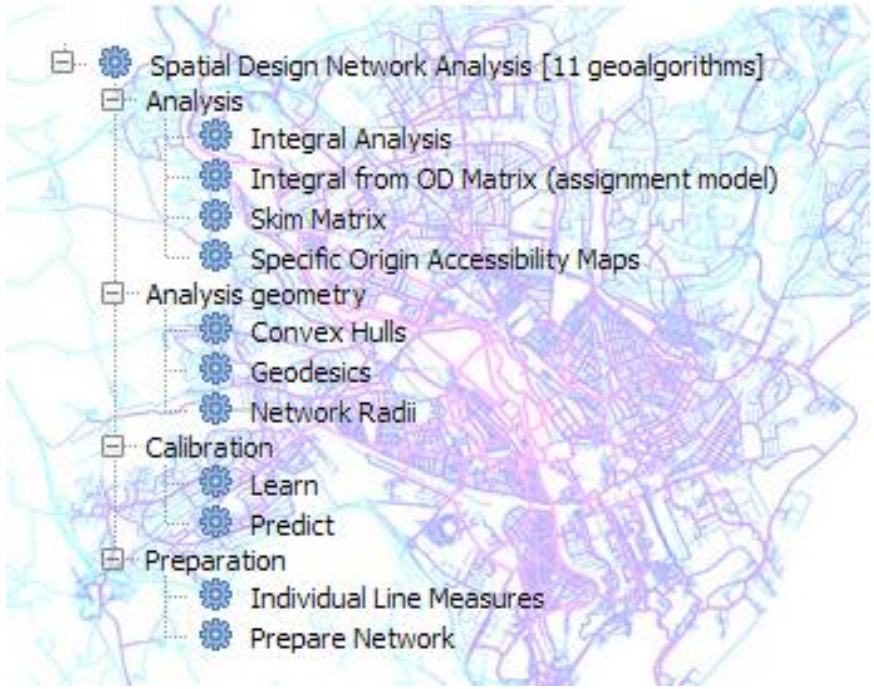
November 2019

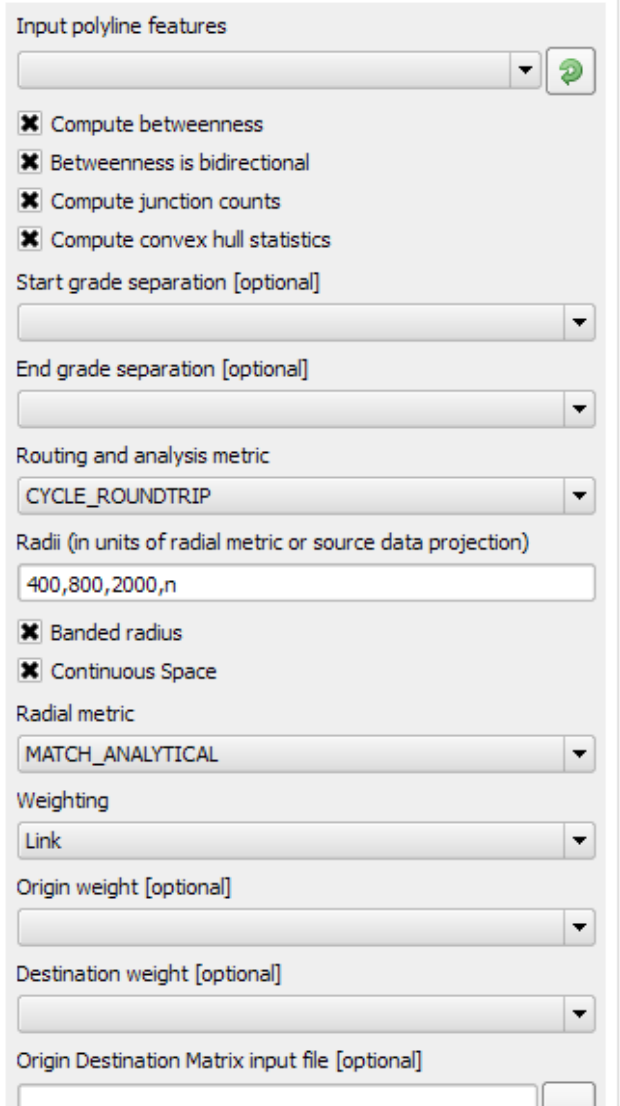
Today's talk

- Theory
  - Networks vs Spatial Networks
  - Distance, Reach, Closeness, Betweenness, Directness, Eigenvalue Centrality
  - What is accessibility?
- Applications
  - Transport
    - Cycling
    - Walking and accessibility
  - Community cohesion & health
  - Economics: house price, business rates & productivity (local, national)
  - Defining regions

About the sDNA software

- Software we've produced since 2011
  - Plugin for ArcGIS, QGIS, Autocad
  - Command line tools
  - Python API
  - Standard version = free and open source
  - C++/OpenMP backend, Python frontend

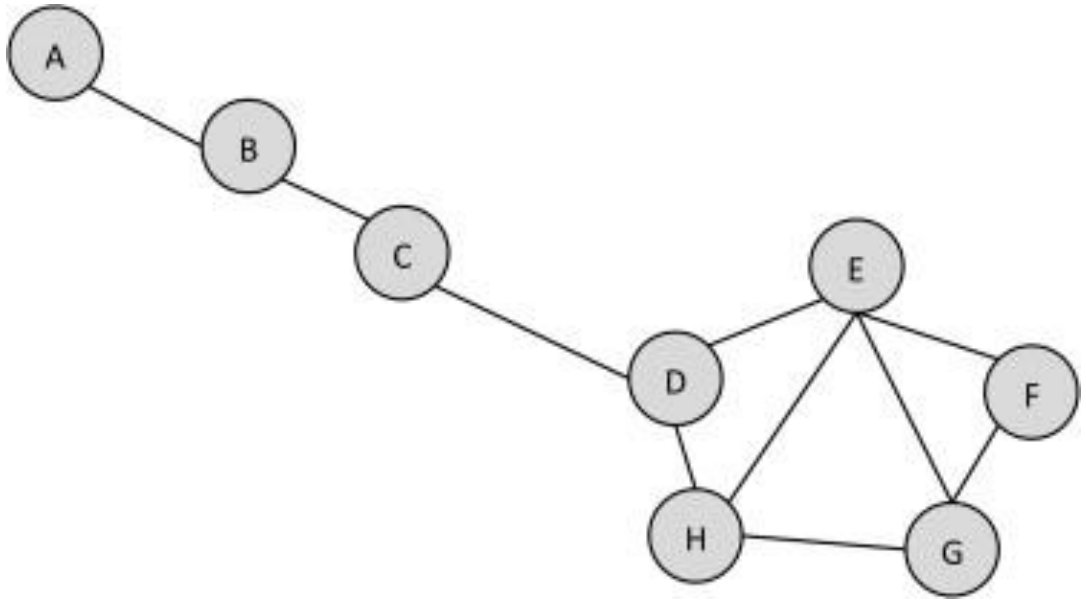




- All examples you are about to see use sDNA

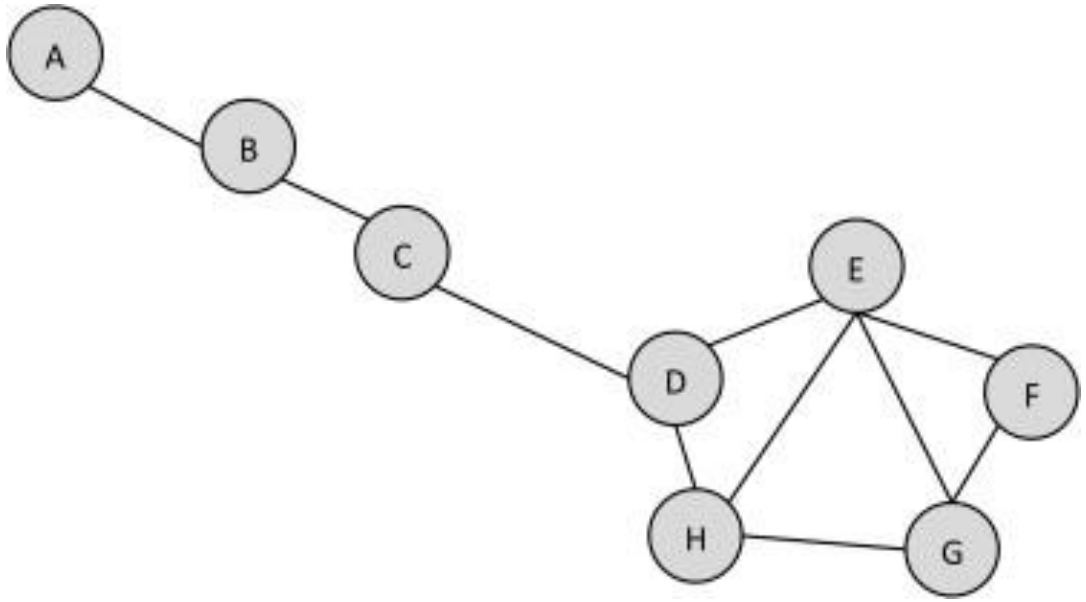
Networks

- Networks
  - Nodes connected by links
  - Measuring distance through network



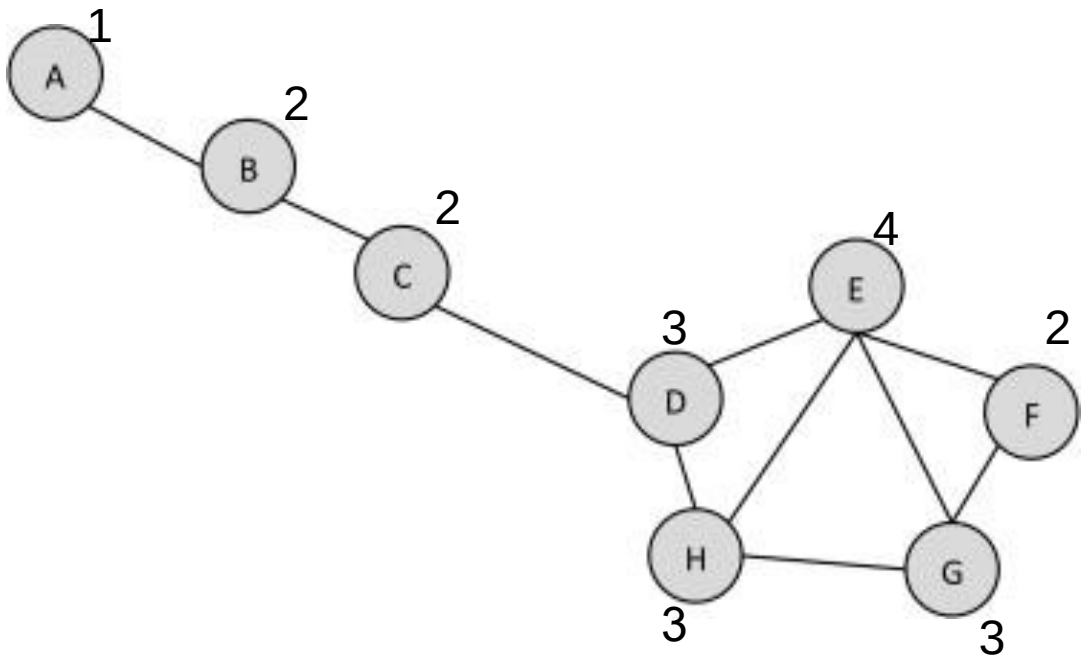
Networks

- Networks
  - Nodes connected by links
  - Centrality measures
    - Degree centrality: number of links from a node
    - Mean distance: mean number of hops from a given node to the others
    - Closeness: 1/mean distance
    - Betweenness: number of shortest paths (everywhere to everywhere) that pass through a given node
    - Reach: number of nodes within given distance



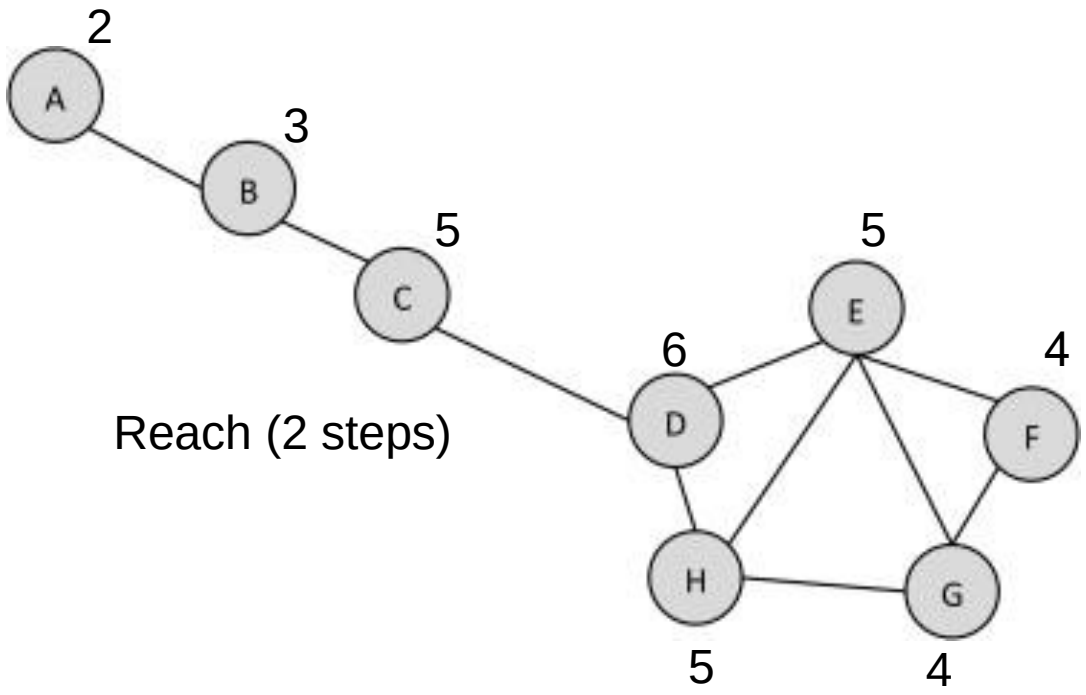
Networks

- Networks
  - Nodes connected by links
  - Centrality measures
    - Degree centrality: number of links from a node
    - Mean distance: mean number of hops from a given node to the others
    - Closeness: 1/mean distance
    - Betweenness: number of shortest paths (everywhere to everywhere) that pass through a given node
    - Reach: number of nodes within given distance



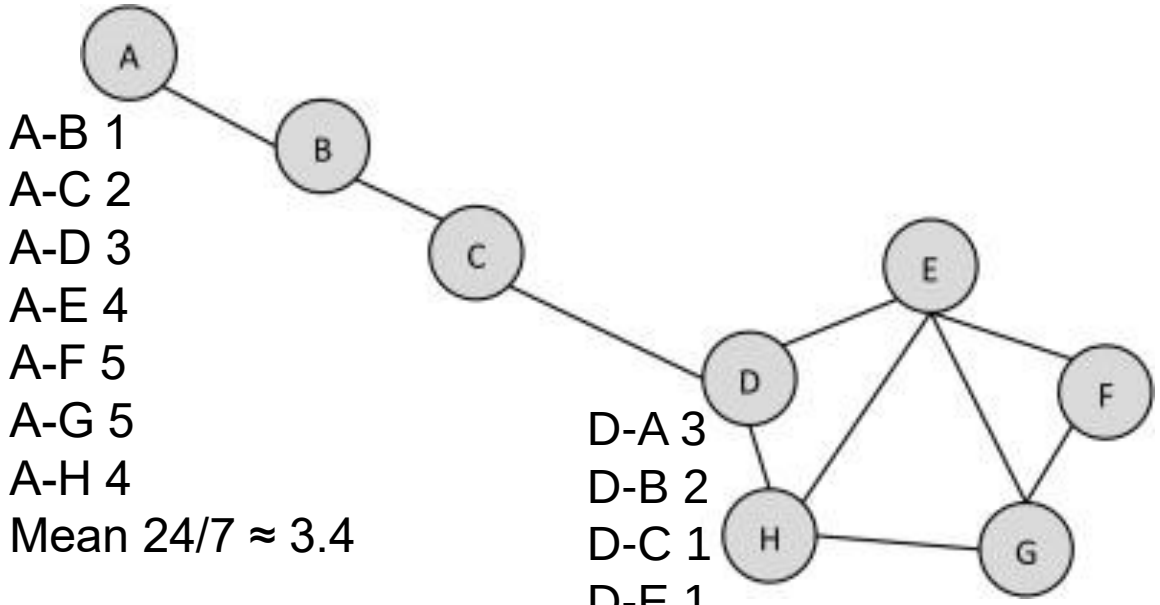
Networks

- Networks
  - Nodes connected by links
  - Centrality measures
    - Degree centrality: number of links from a node
    - Mean distance: mean number of hops from a given node to the others
    - Closeness: 1/mean distance
    - Betweenness: number of shortest paths (everywhere to everywhere) that pass through a given node
    - Reach: number of nodes within given distance



Networks

- Networks
  - Nodes connected by links
  - Centrality measures
    - Degree centrality: number of links from a node
    - Mean distance: mean number of hops from a given node to the others
    - Closeness: 1/mean distance
    - Betweenness: number of shortest paths (everywhere to everywhere) that pass through a given node
    - Reach: number of nodes within given distance

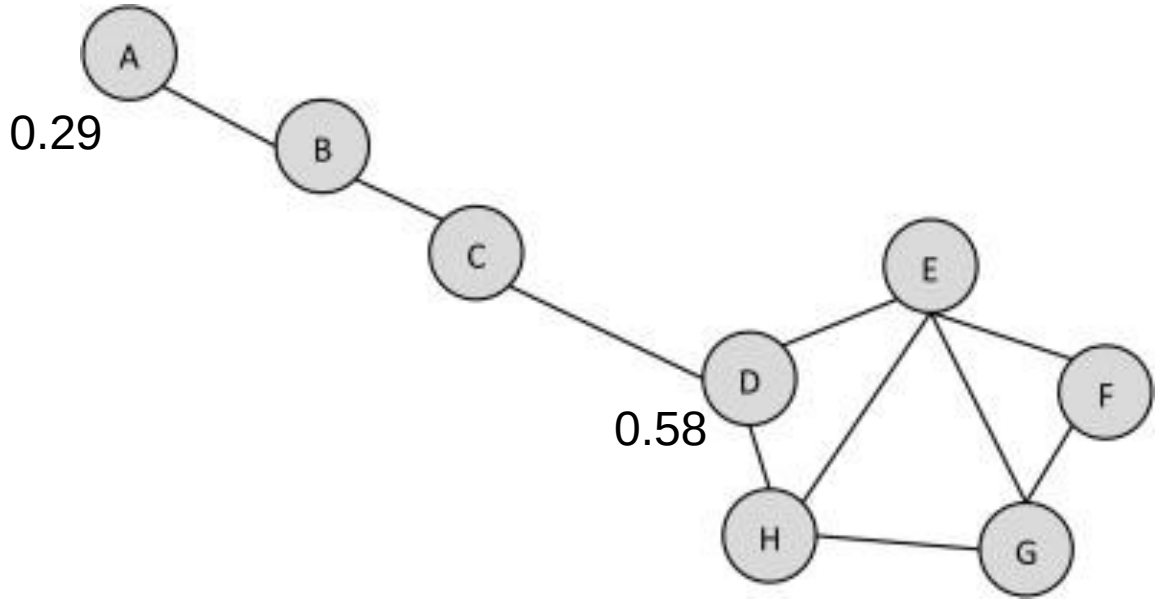


A-B 1
A-C 2
A-D 3
A-E 4
A-F 5
A-G 5
A-H 4
Mean 24/7 ≈ 3.4

D-A 3
D-B 2
D-C 1
D-E 1
D-F 2
D-G 2
D-H 1
Mean 12/7 ≈ 1.7

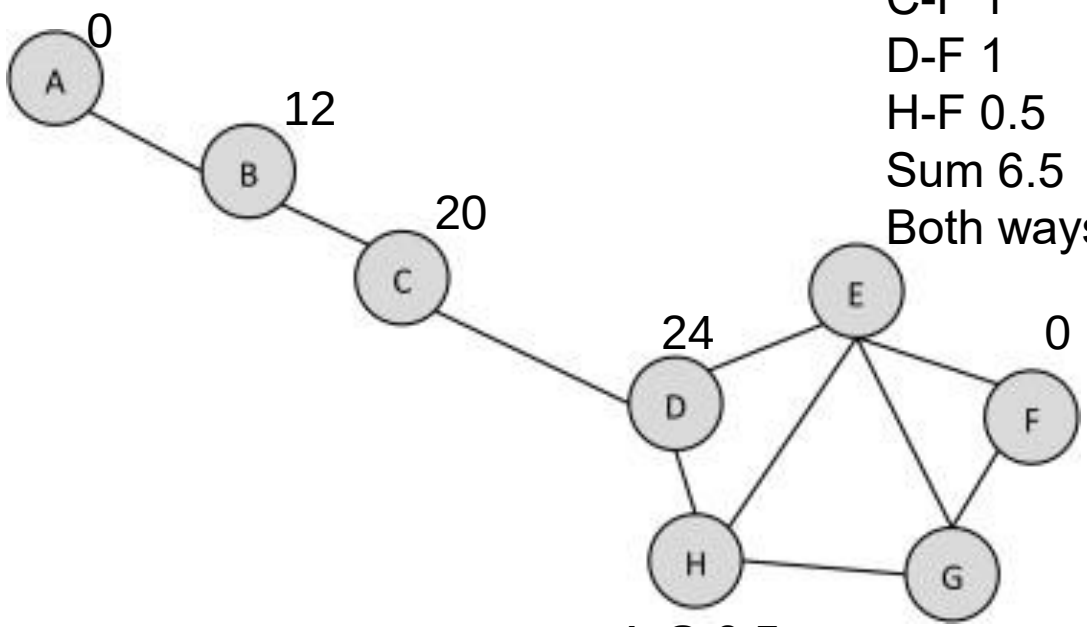
Networks

- Networks
  - Nodes connected by links
  - Centrality measures
    - Degree centrality: number of links from a node
    - Mean distance: mean number of hops from a given node to the others
    - Closeness: 1/mean distance
    - Betweenness: number of shortest paths (everywhere to everywhere) that pass through a given node
    - Reach: number of nodes within given distance



Networks

- Networks
  - Nodes connected by links
  - Centrality measures
    - Degree centrality: number of links from a node
    - Mean distance: mean number of hops from a given node to the others
    - Closeness: 1/mean distance
    - Betweenness: number of shortest paths (everywhere to everywhere) that pass through a given node
    - Reach: number of nodes within given distance



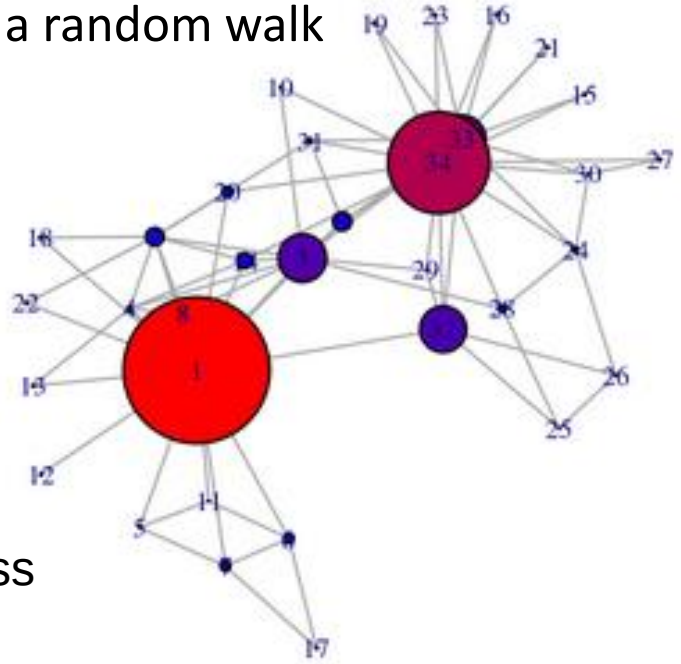
A-G 0.5
B-G 0.5
C-G 0.5
D-G 0.5
A-F 1
B-F 1
C-F 1
D-F 1
H-F 0.5
Sum 6.5
Both ways 13

A-G 0.5
B-G 0.5
C-G 0.5
D-G 0.5
Sum 2
Both ways 4

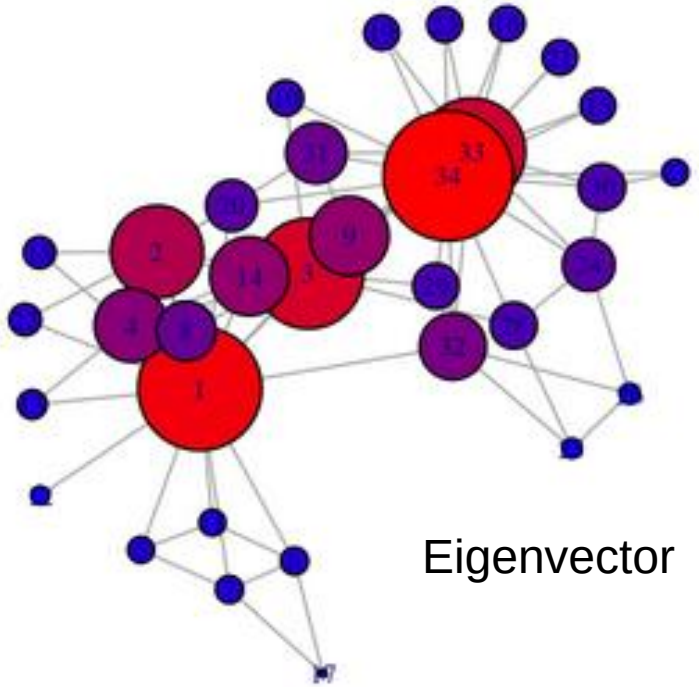
H-F 0.5
Both ways 1

Networks

- Eigenvector centrality
  - Principal eigenvector of connectivity matrix
  - Google PageRank
  - Betweenness of a random walk



Betweenness



Eigenvector

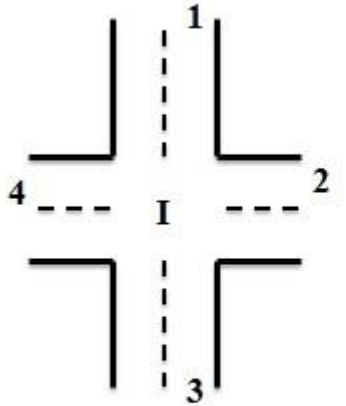
http://www.rpubs.com/shestakoff/sna_lab4

What's different about Spatial Networks?

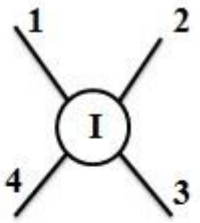
- Ontological
  - Nodes have positions in space
  - Links have shape: they may not be straight lines
- Analytical
  - We care more about links than nodes => dual representation
  - We can use spatial characteristics to define weightings and distance
  - We can restrict measures to spatial locality
  - We can use spatially explicit measures (more on this later)

Dual representation

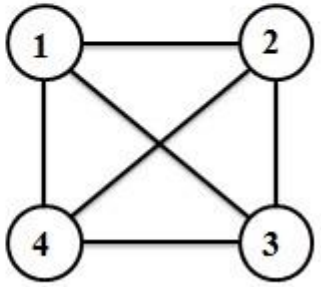
- Links become nodes,
  nodes become links (Añez 1996)
- Nodes (e.g. roads) now have spatial extent
  - Self closeness
  - Self betweenness (SoftwareX? 2019)



Road Network with an intersection



Primal Graph

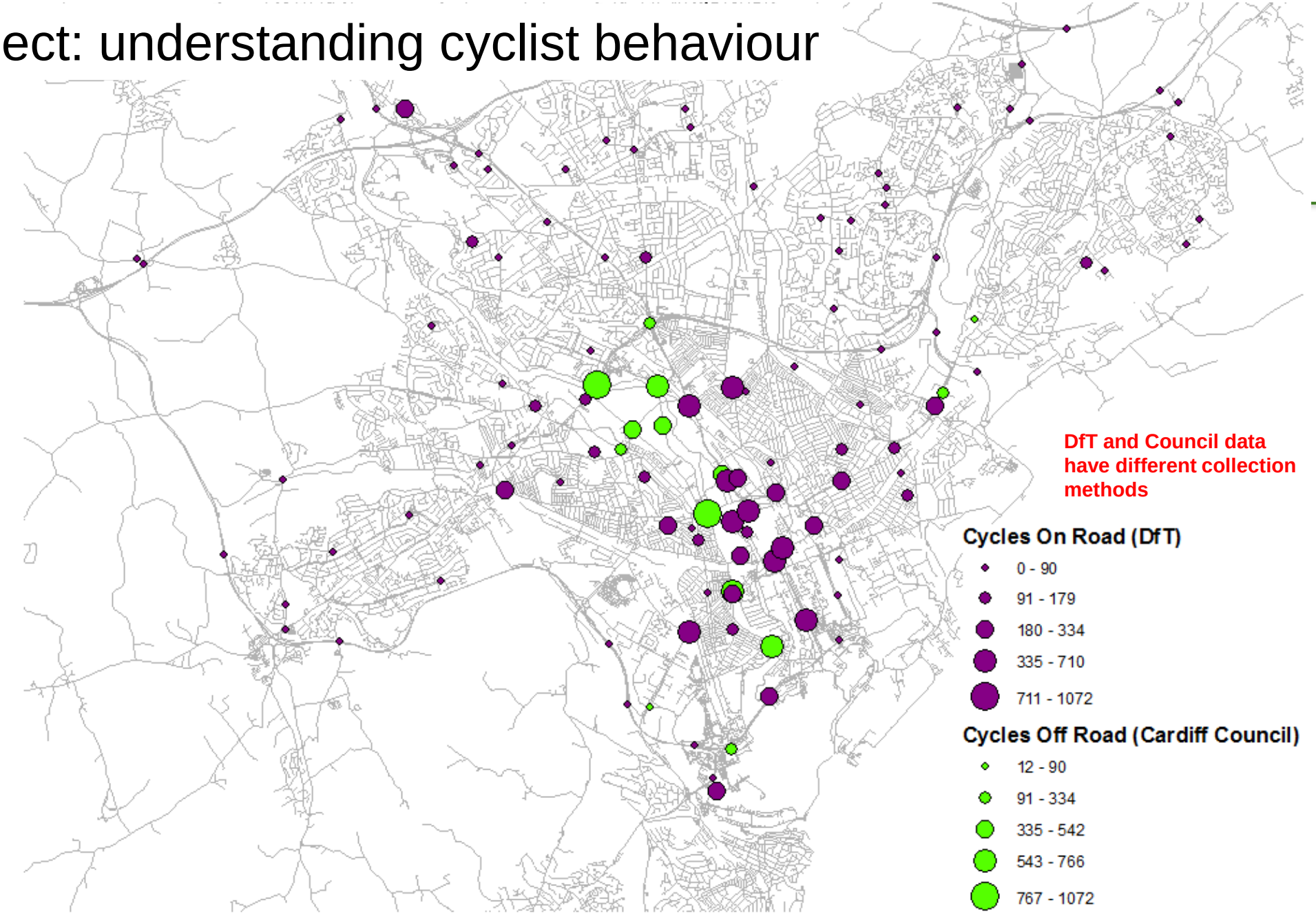


Dual Graph

Using space and localizing measures

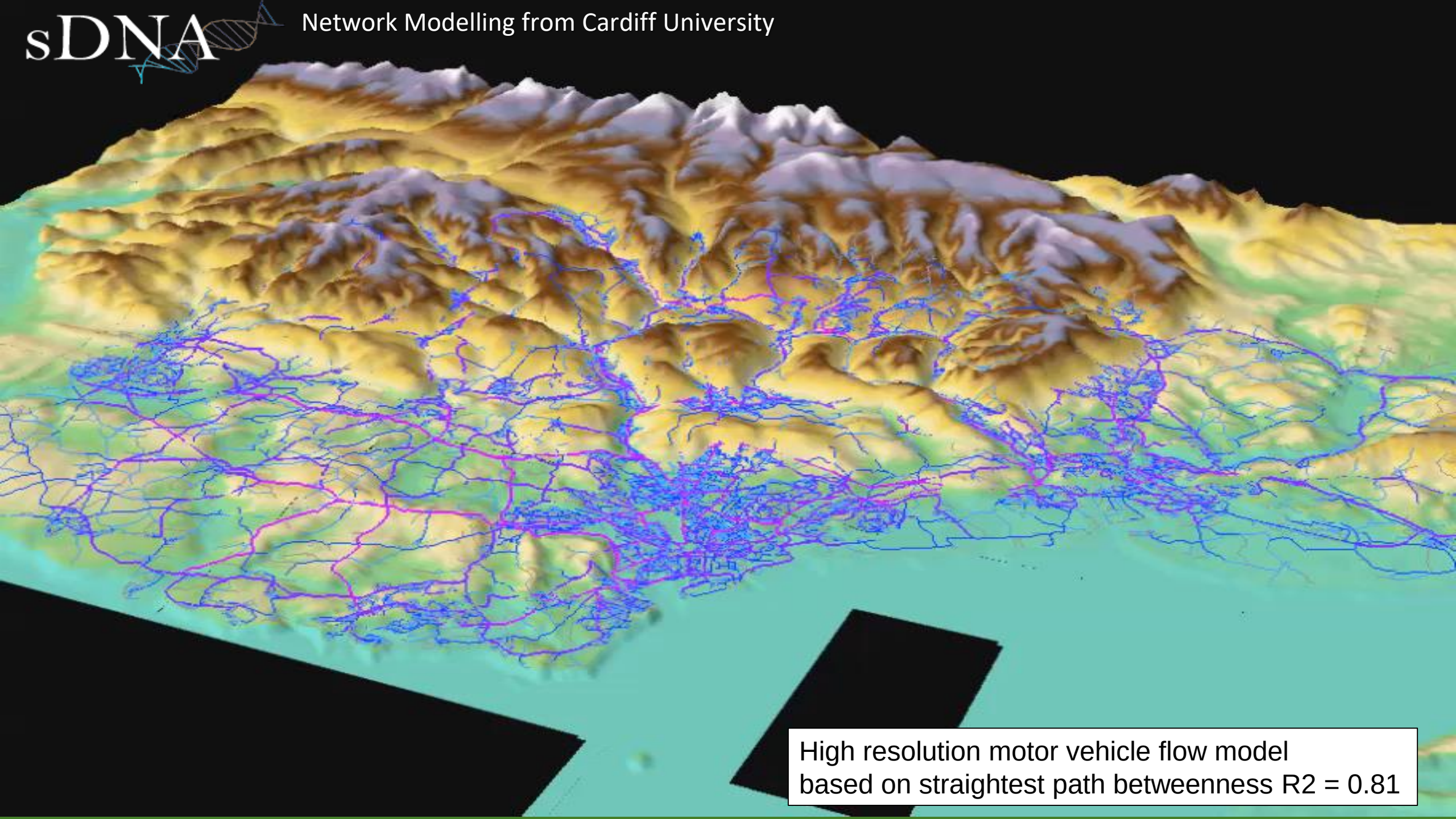
- Distance metrics
  - Length along network: network-Euclidean i.e. shortest distance
  - Angular: least turning, most direct
  - Topological: fewest junctions
  - Travel time
  - Cyclist distance: negative utility including aversion to hills and traffic
  - Pedestrian distance: blend of shortest plus straightest with randomization
- Localizing
  - “Cut out” the network surrounding each link within e.g. a 5km network-Euclidean buffer
  - Compute network statistics for the locality
  - Details
    - Only remove origins/destinations or also remove links for routing? How to handle paths that exceed size of the locality? Does it matter – compute time, desired results? (IJGIS 2015)

Project: understanding cyclist behaviour

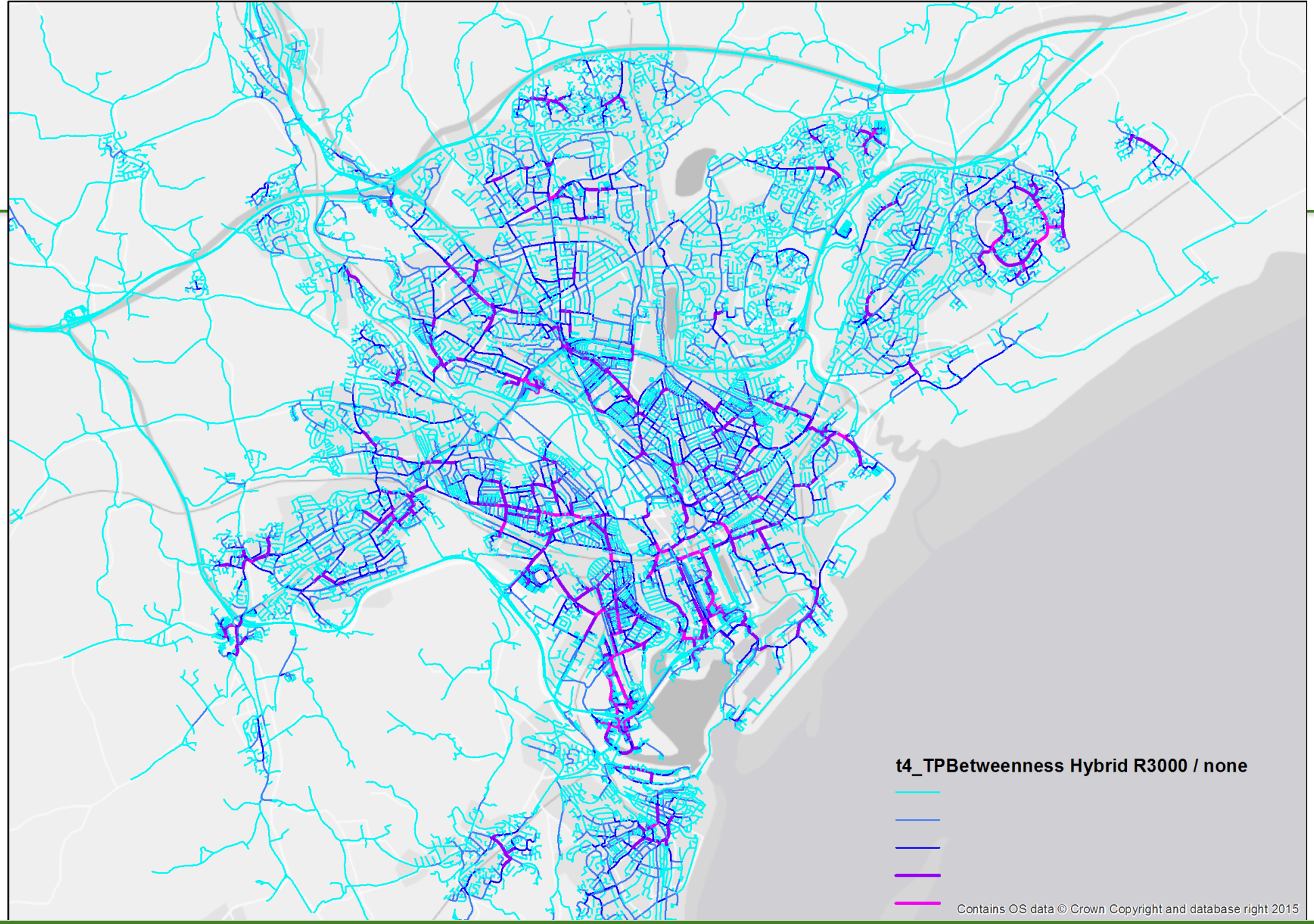


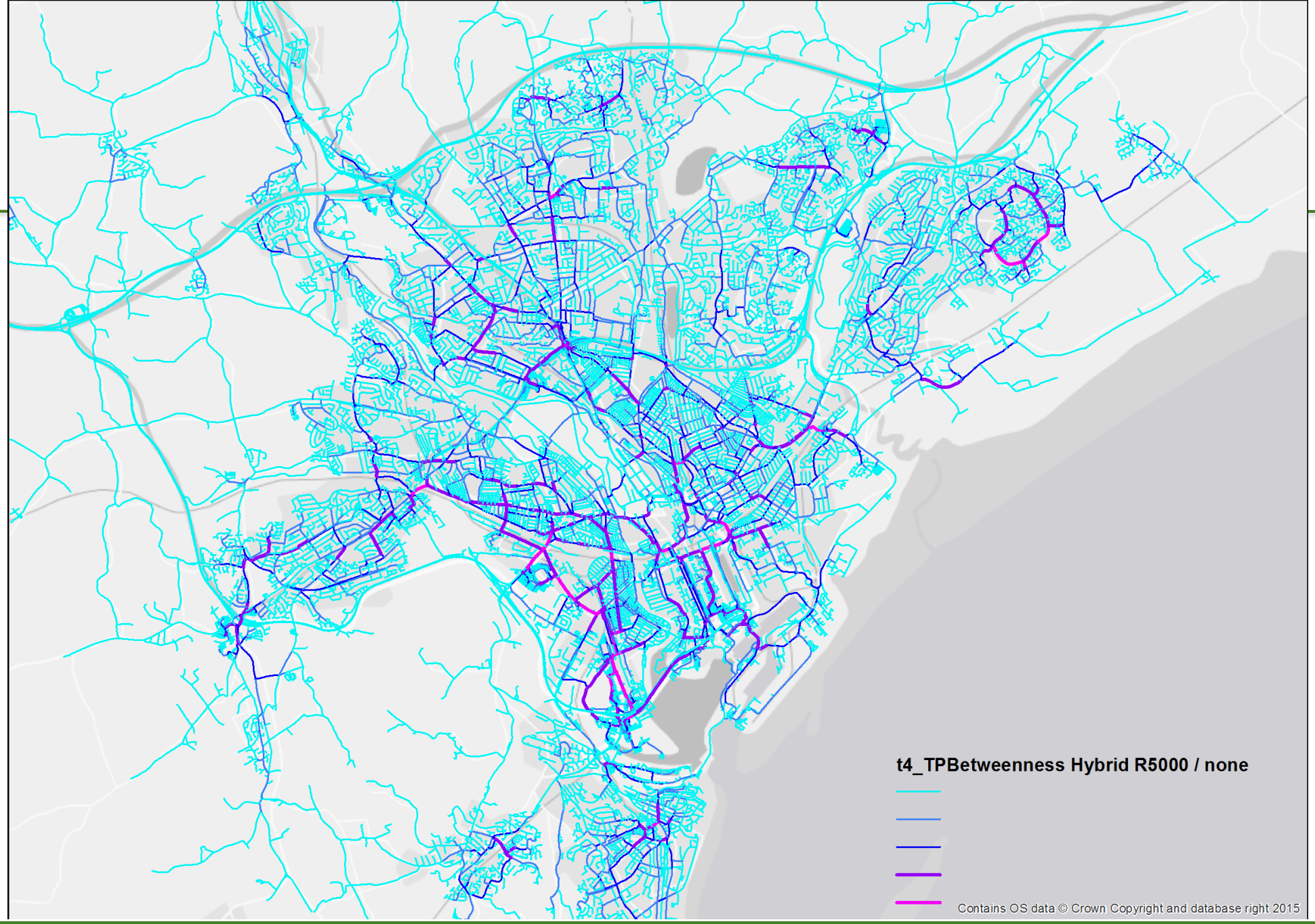
sDNA

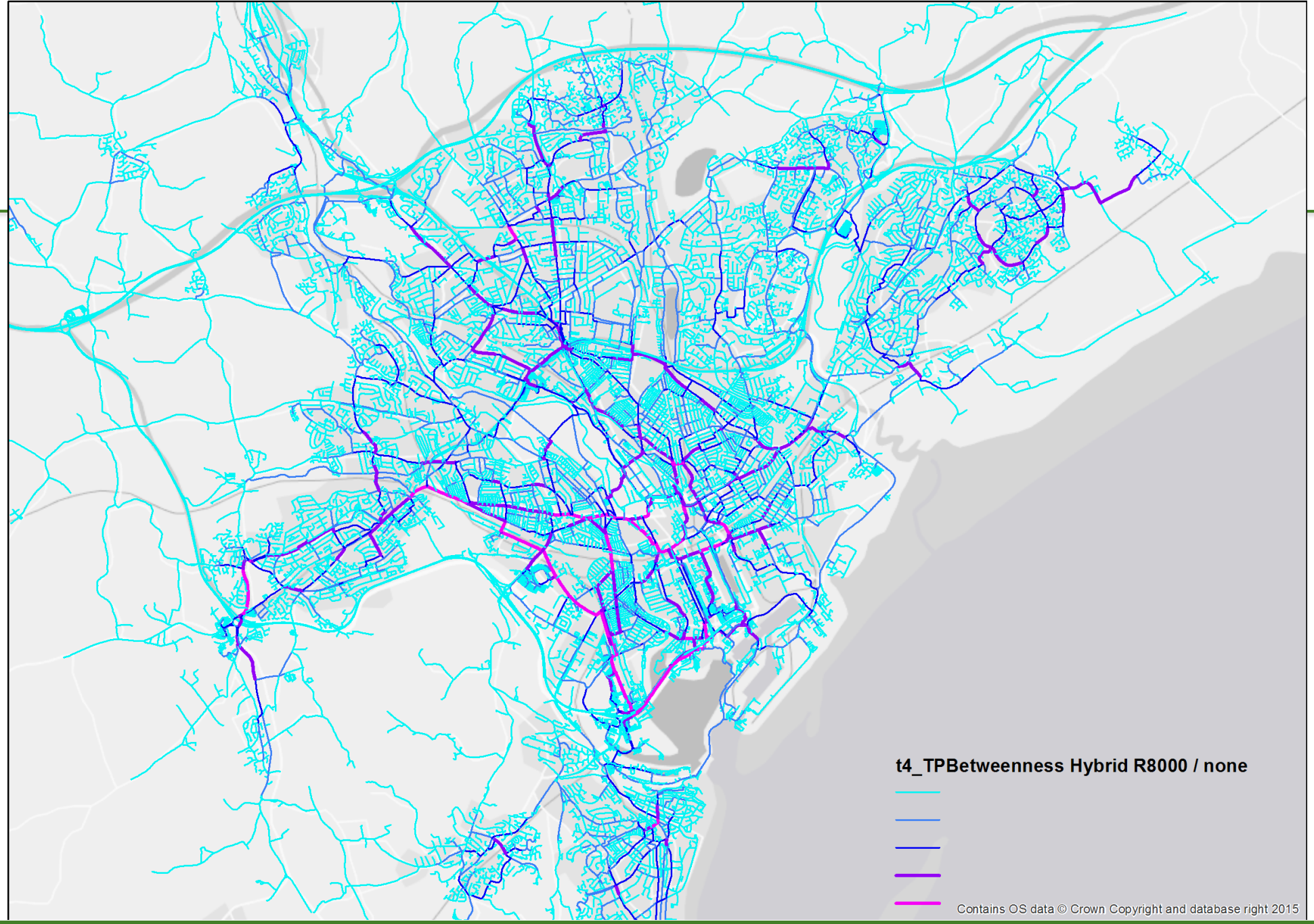
Network Modelling from Cardiff University

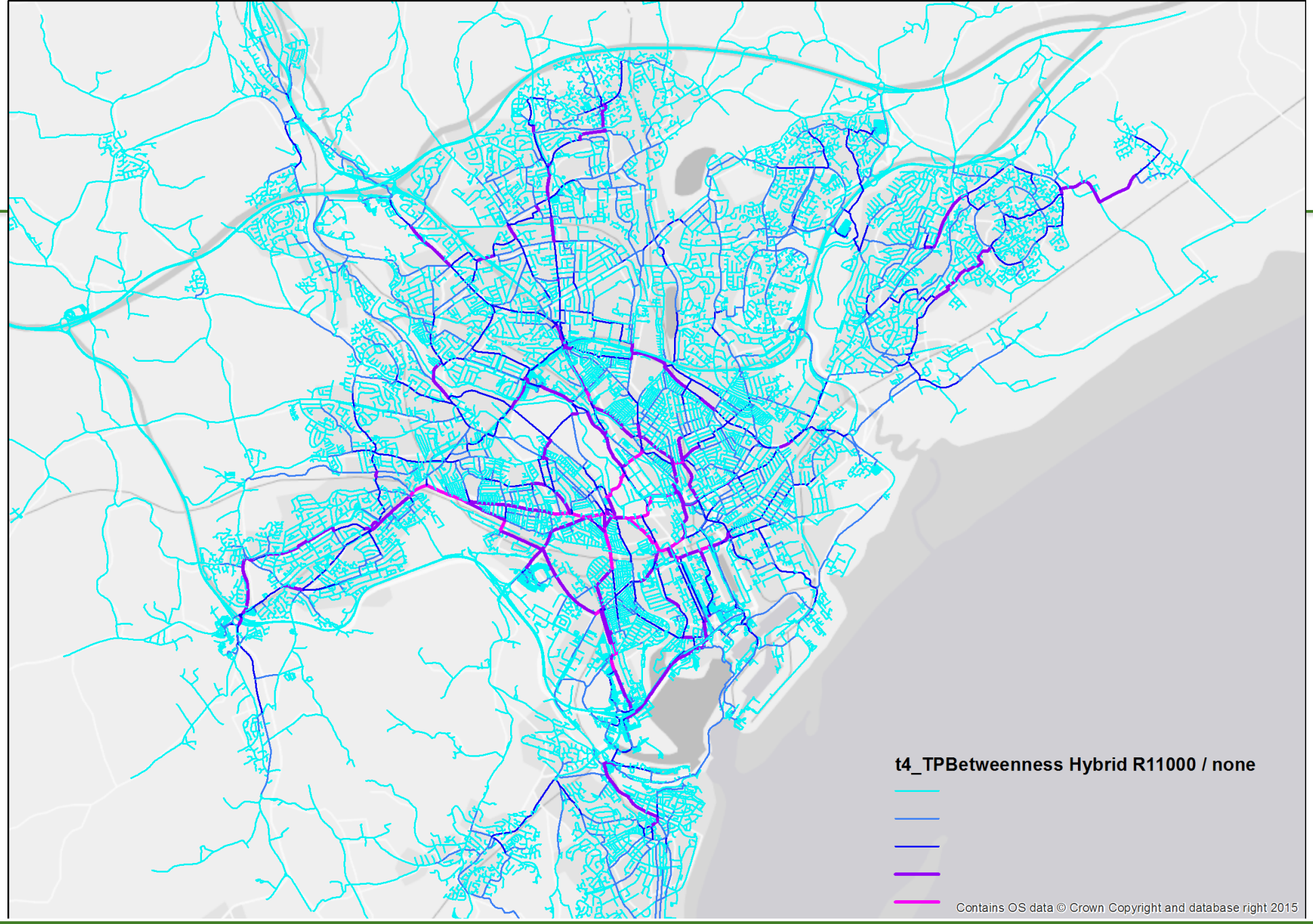


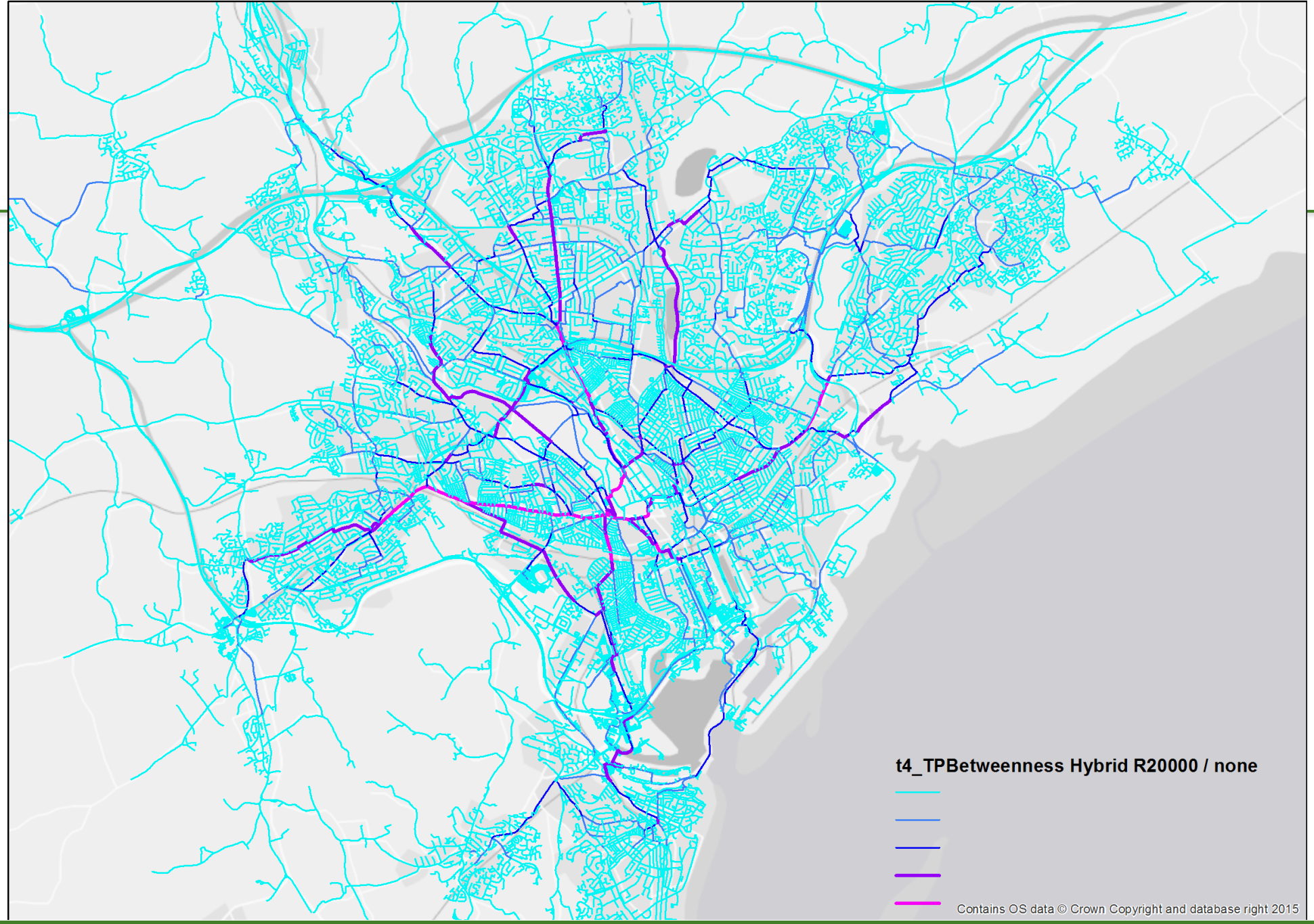
High resolution motor vehicle flow model
based on straightest path betweenness R2 = 0.81

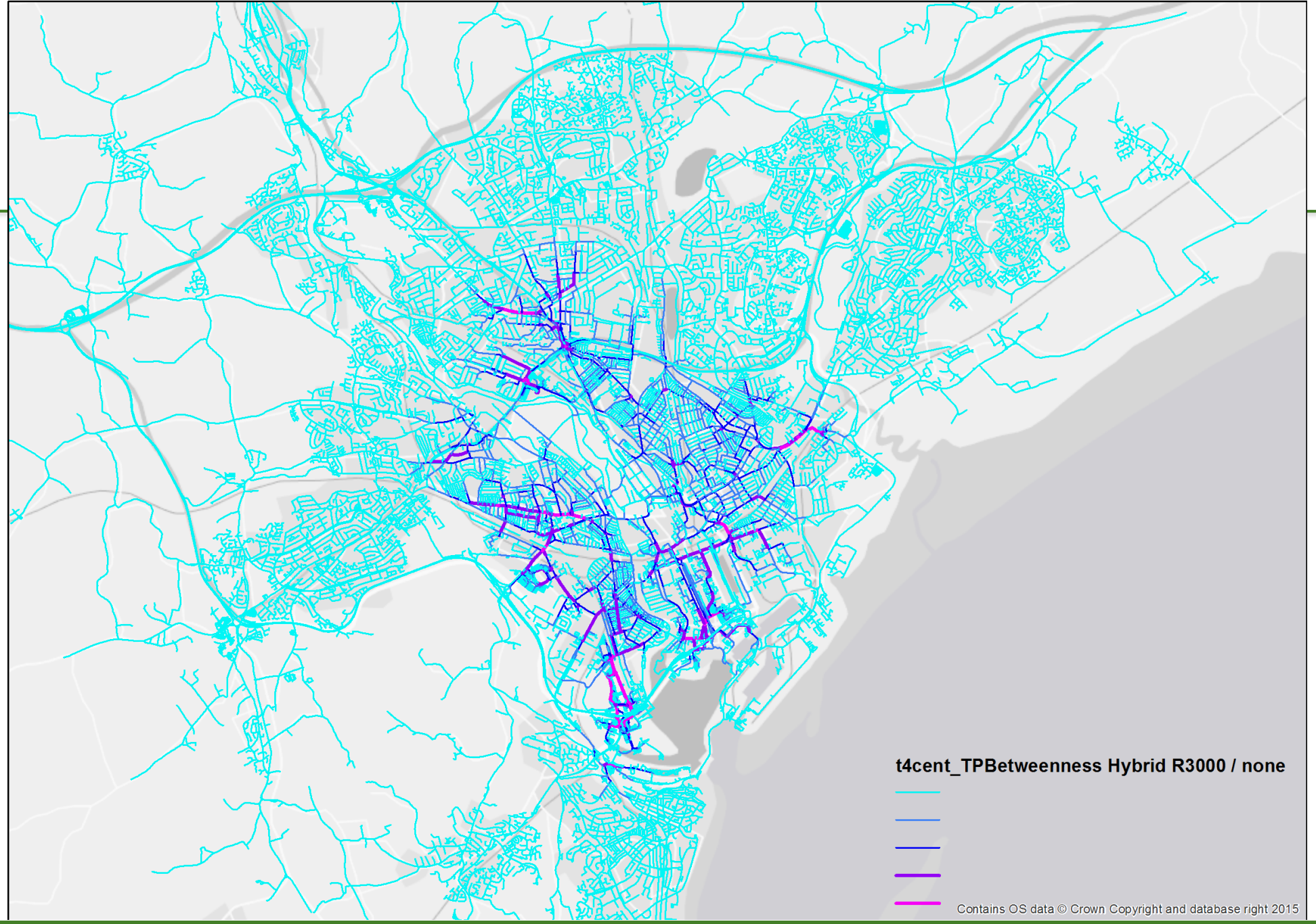

t4_TPBetweenness Hybrid R3000 / none
Contains OS data © Crown Copyright and database right 2015

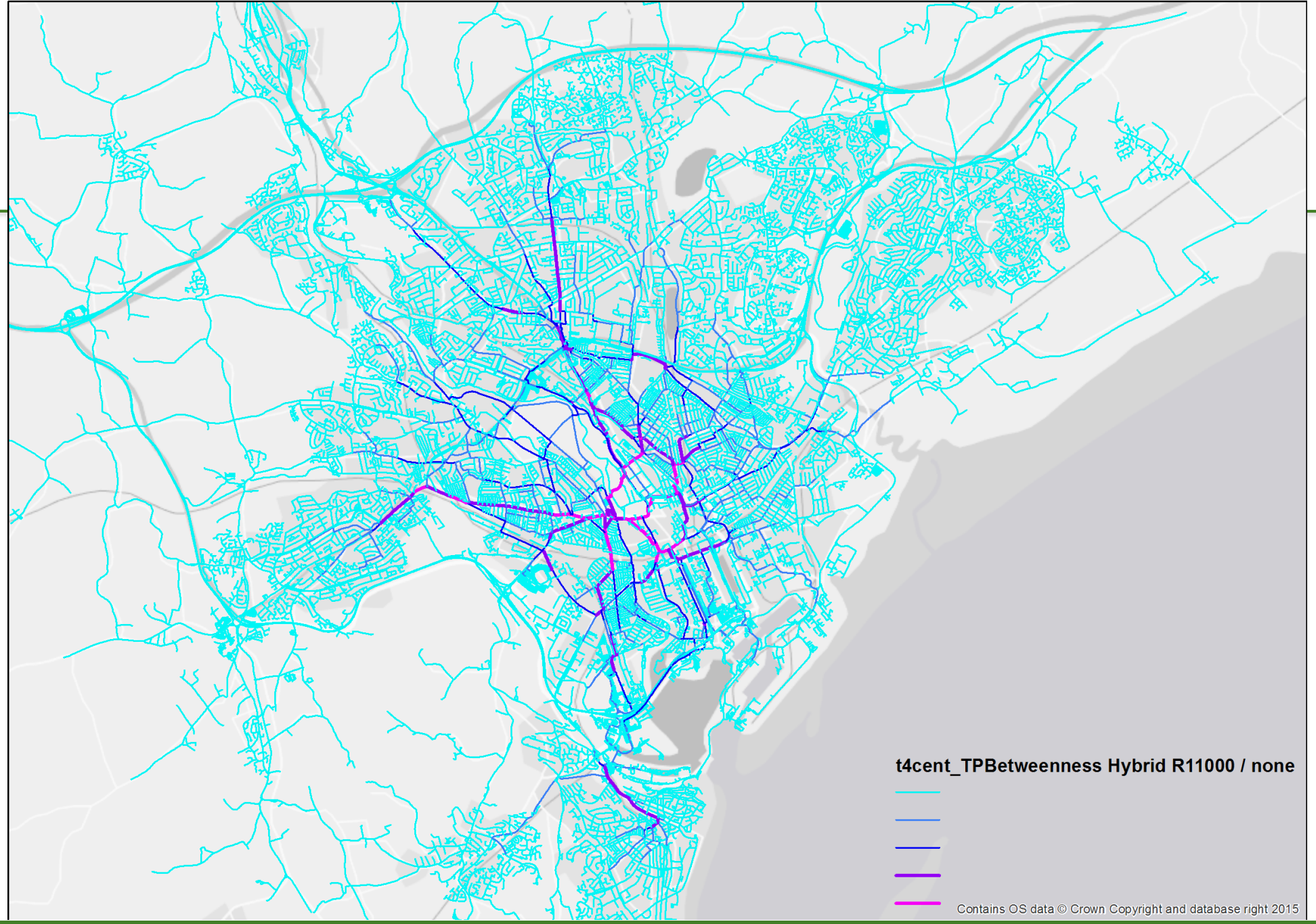

t4_TPBetweenness Hybrid R5000 / none
Contains OS data © Crown Copyright and database right 2015

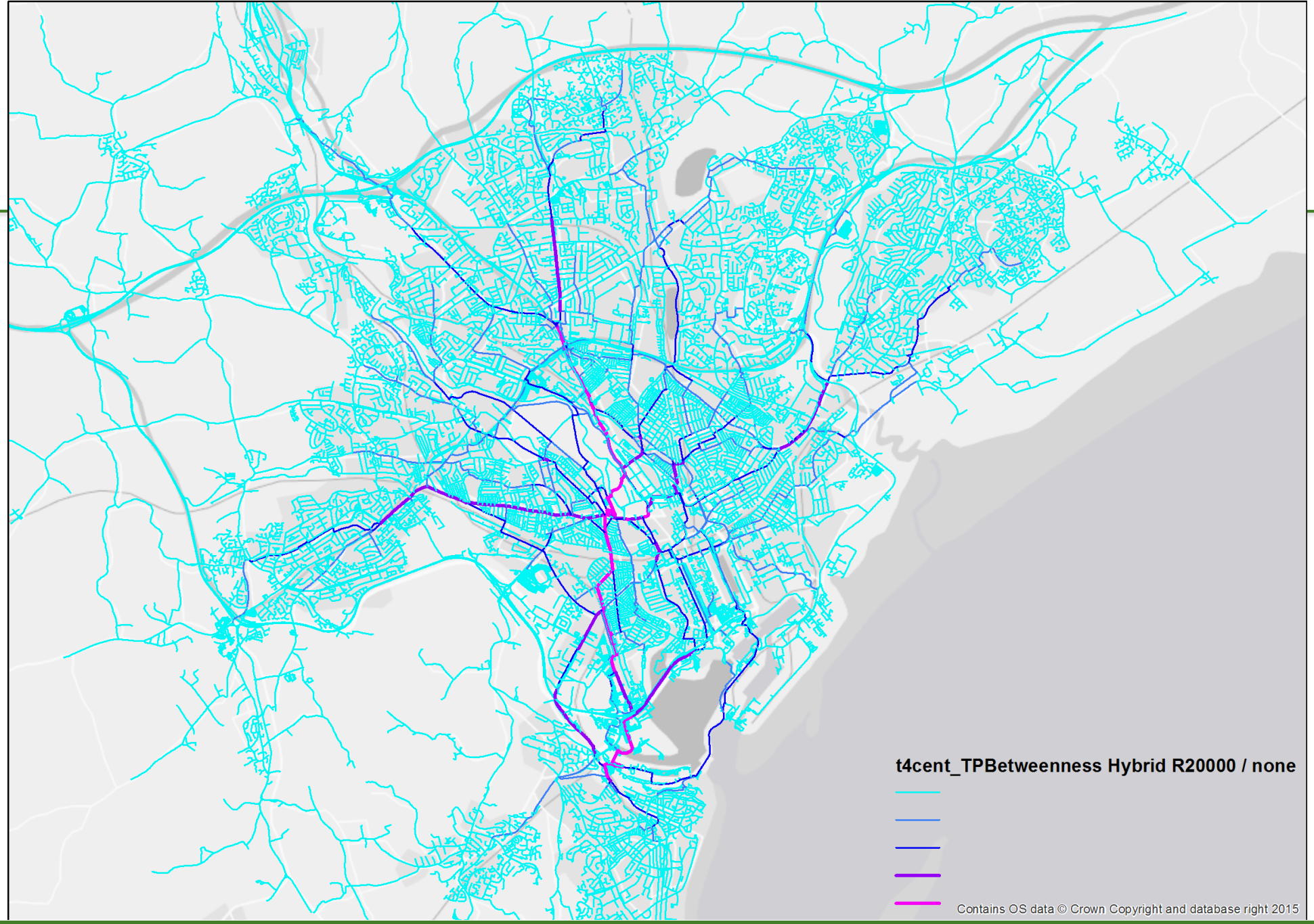

t4_TPBetweenness Hybrid R8000 / none
Contains OS data © Crown Copyright and database right 2015

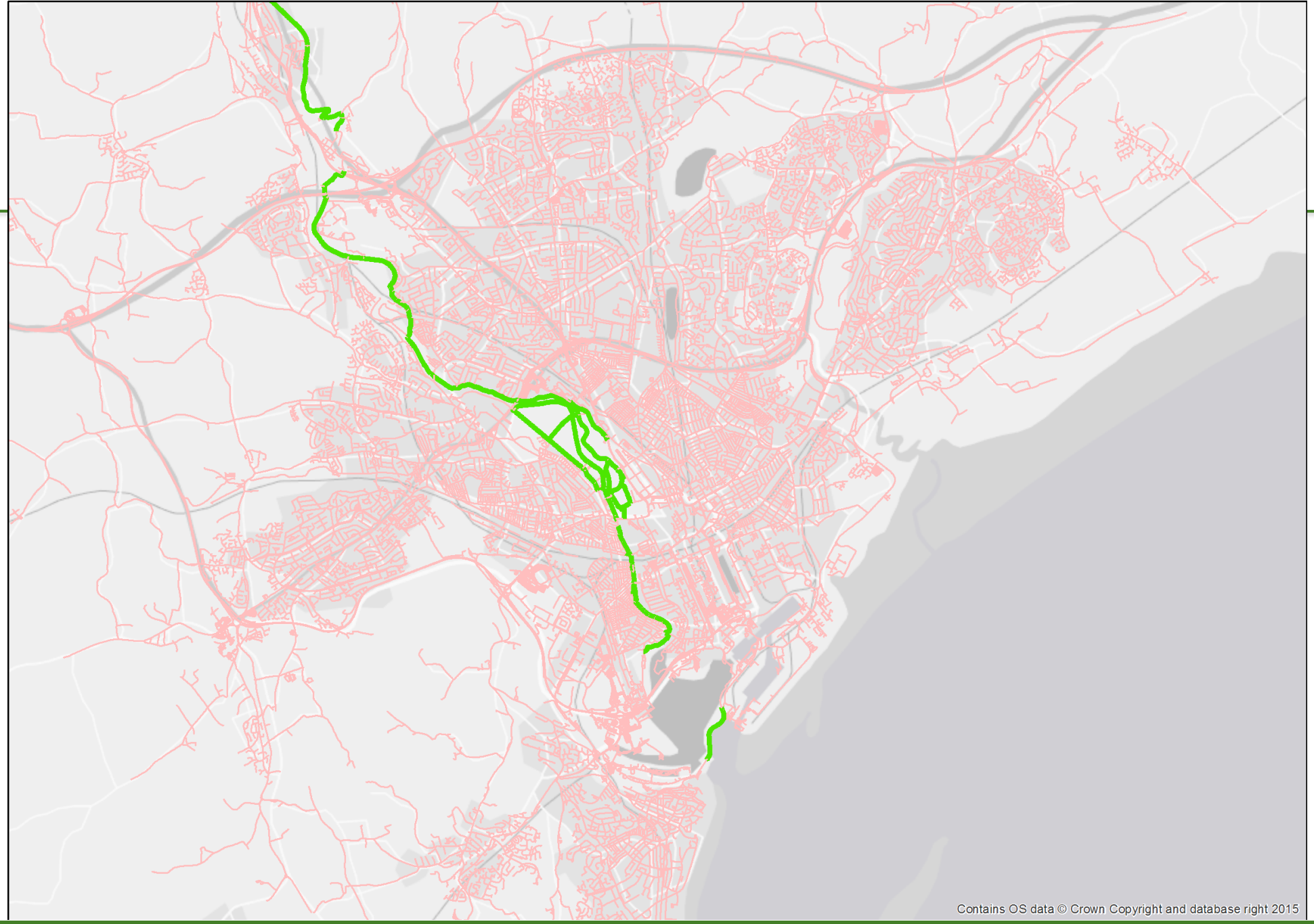

t4_TPBetweenness Hybrid R11000 / none
Contains OS data © Crown Copyright and database right 2015

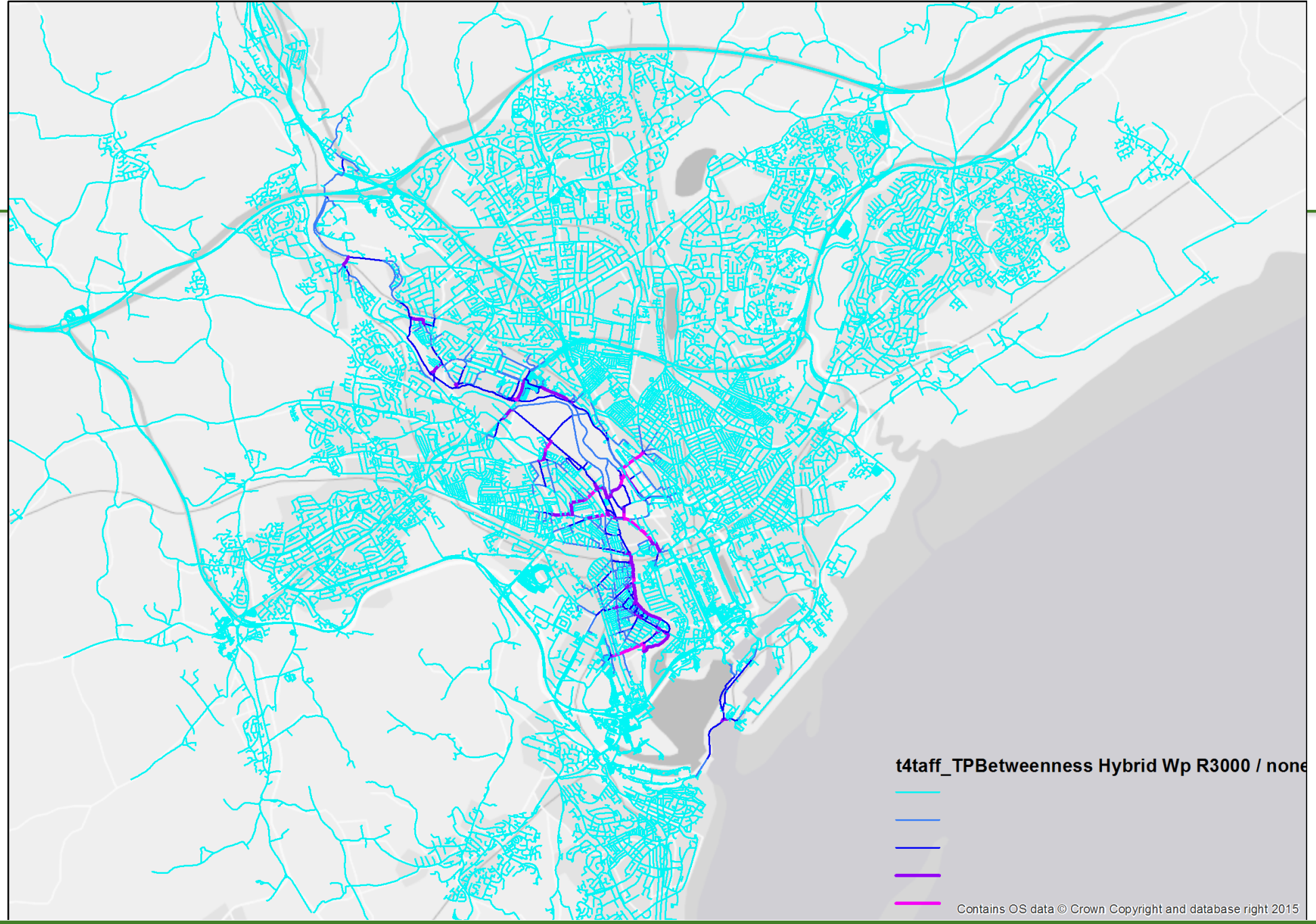

t4_TPBetweenness Hybrid R20000 / none
Contains OS data © Crown Copyright and database right 2015

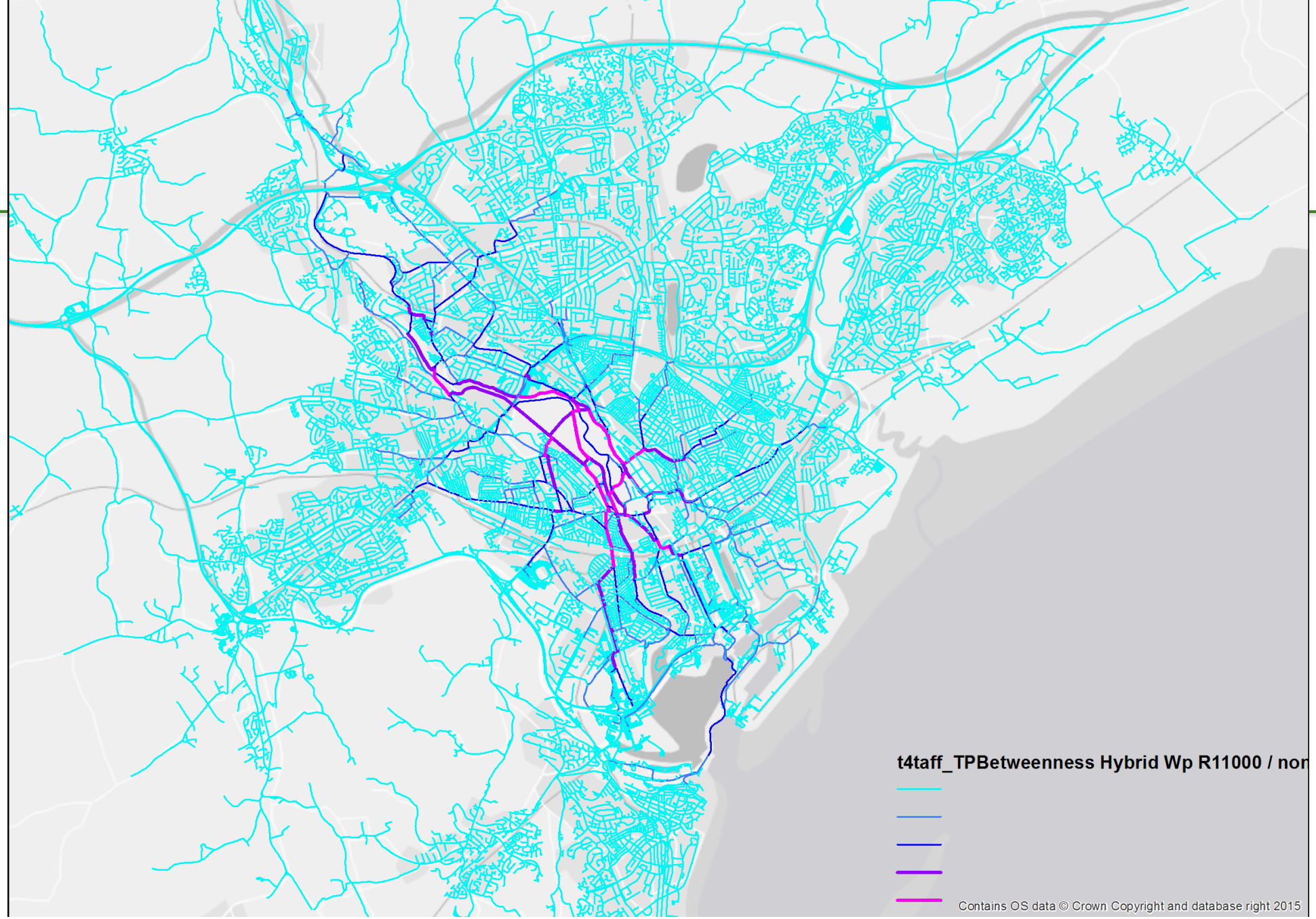

t4cent_TPBetweenness Hybrid R3000 / none
Contains OS data © Crown Copyright and database right 2015

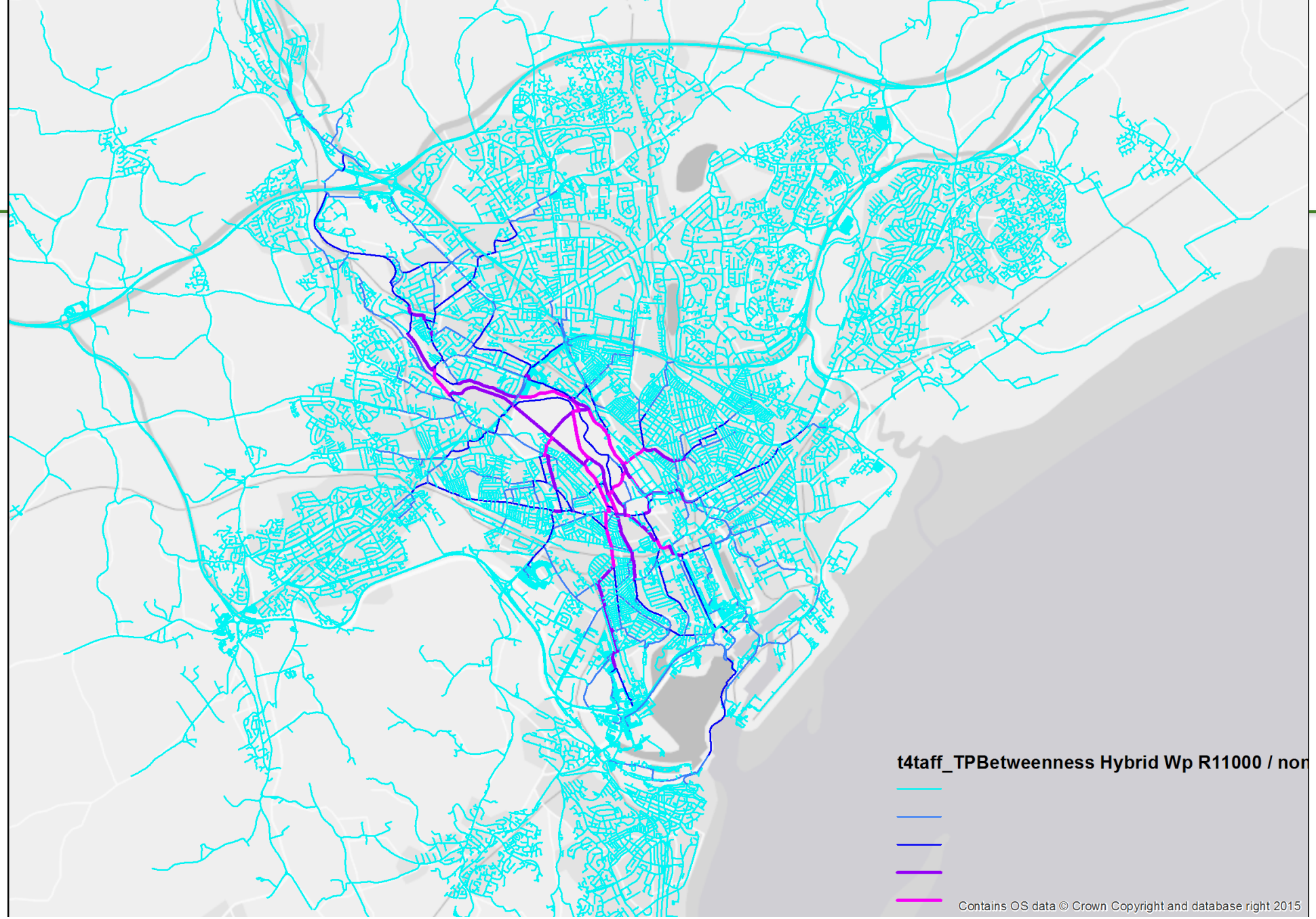

t4cent_TPBetweenness Hybrid R11000 / none
Contains OS data © Crown Copyright and database right 2015

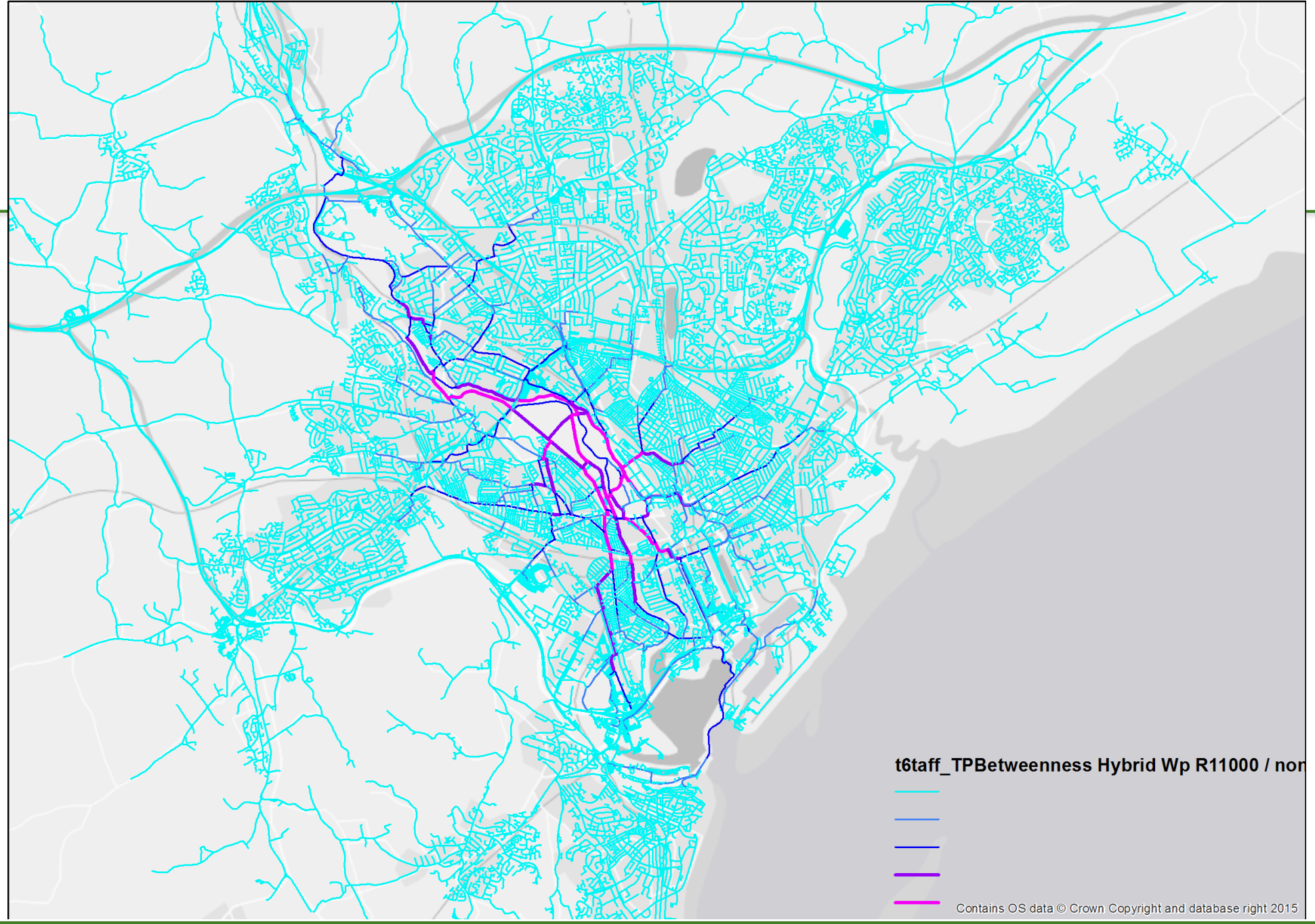

t4cent_TPBetweenness Hybrid R20000 / none
Contains OS data © Crown Copyright and database right 2015

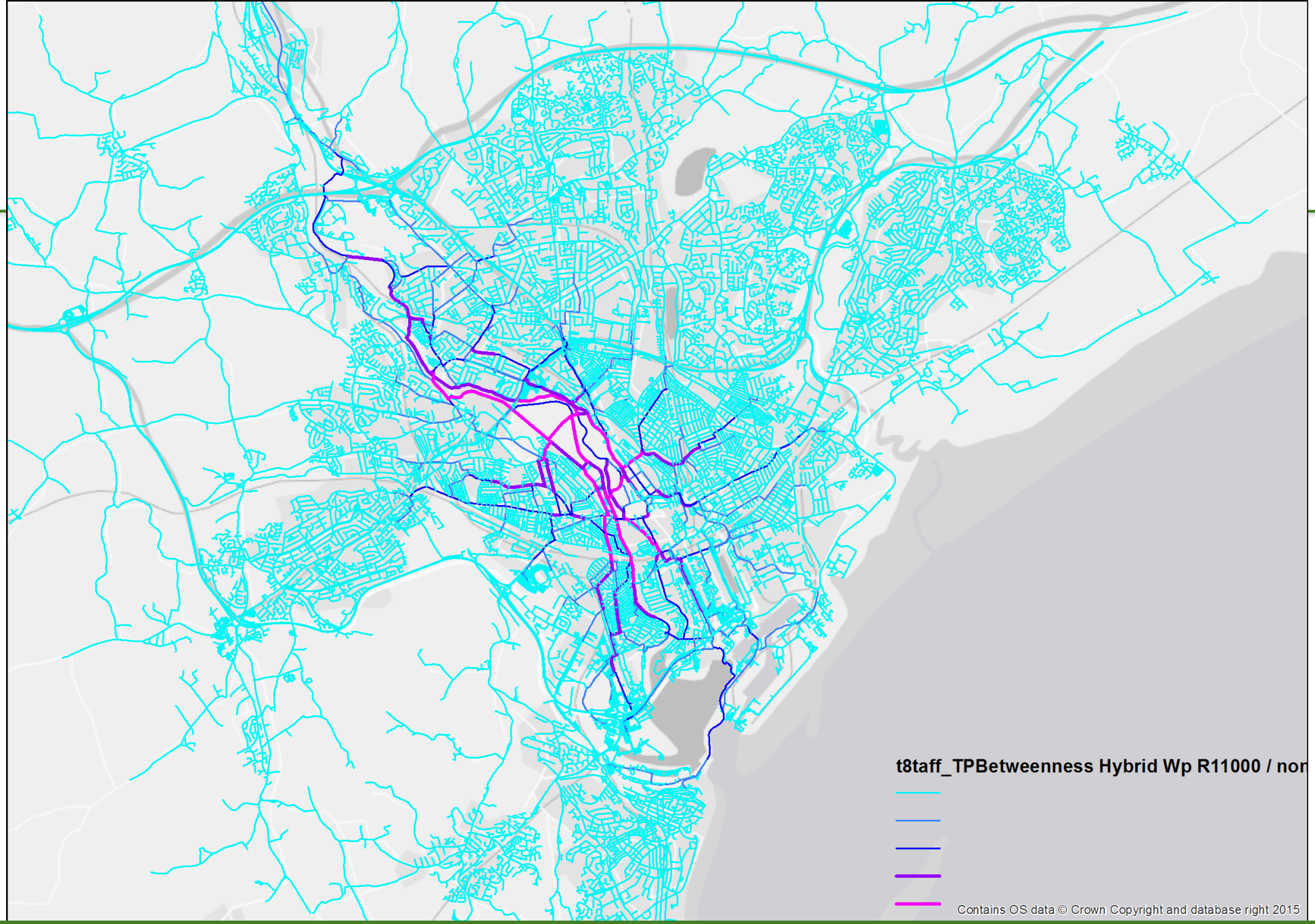

Contains OS data © Crown Copyright and database right 2015

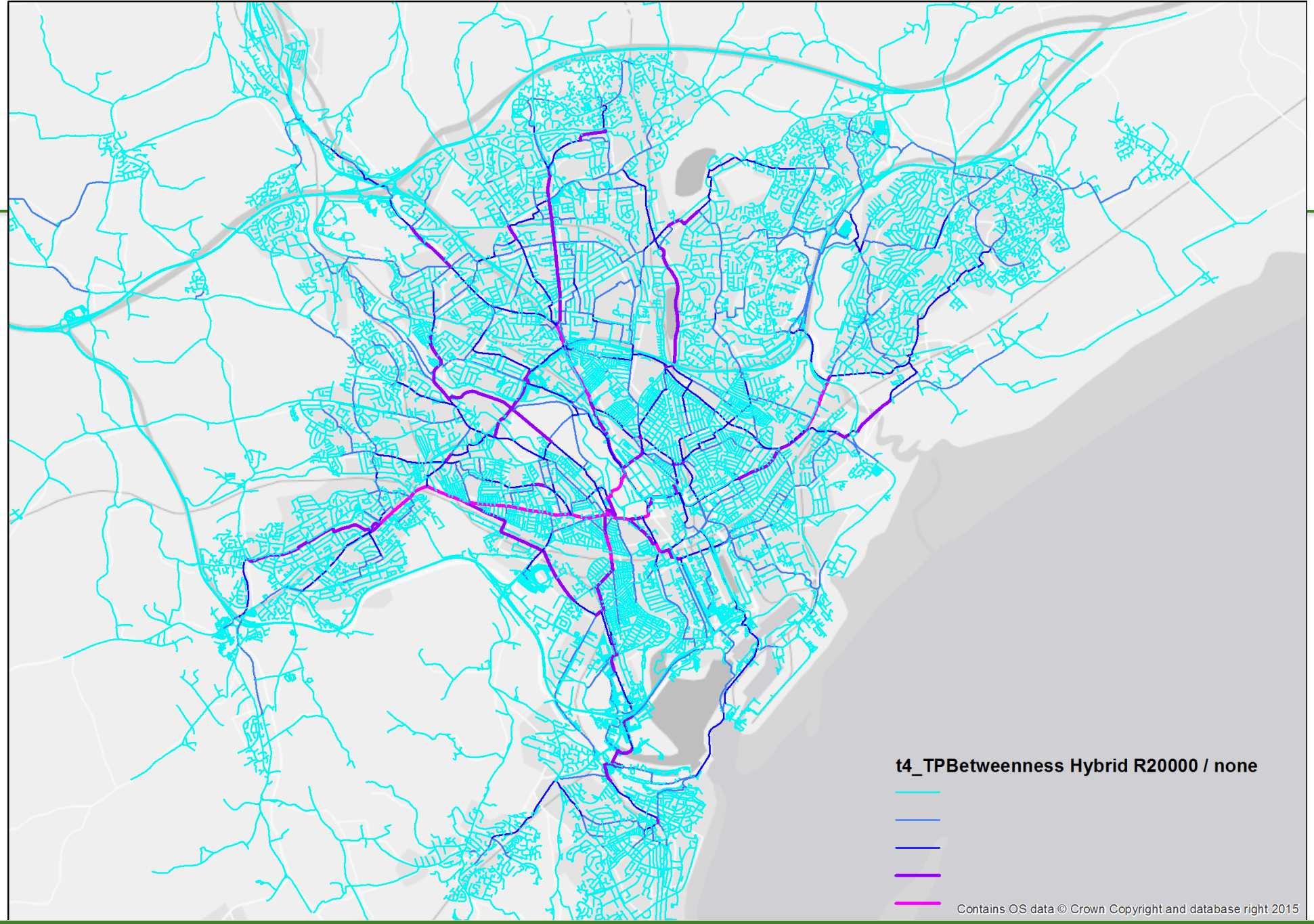

t4taff_TPBetweenness Hybrid Wp R3000 / none
Contains OS data © Crown Copyright and database right 2015

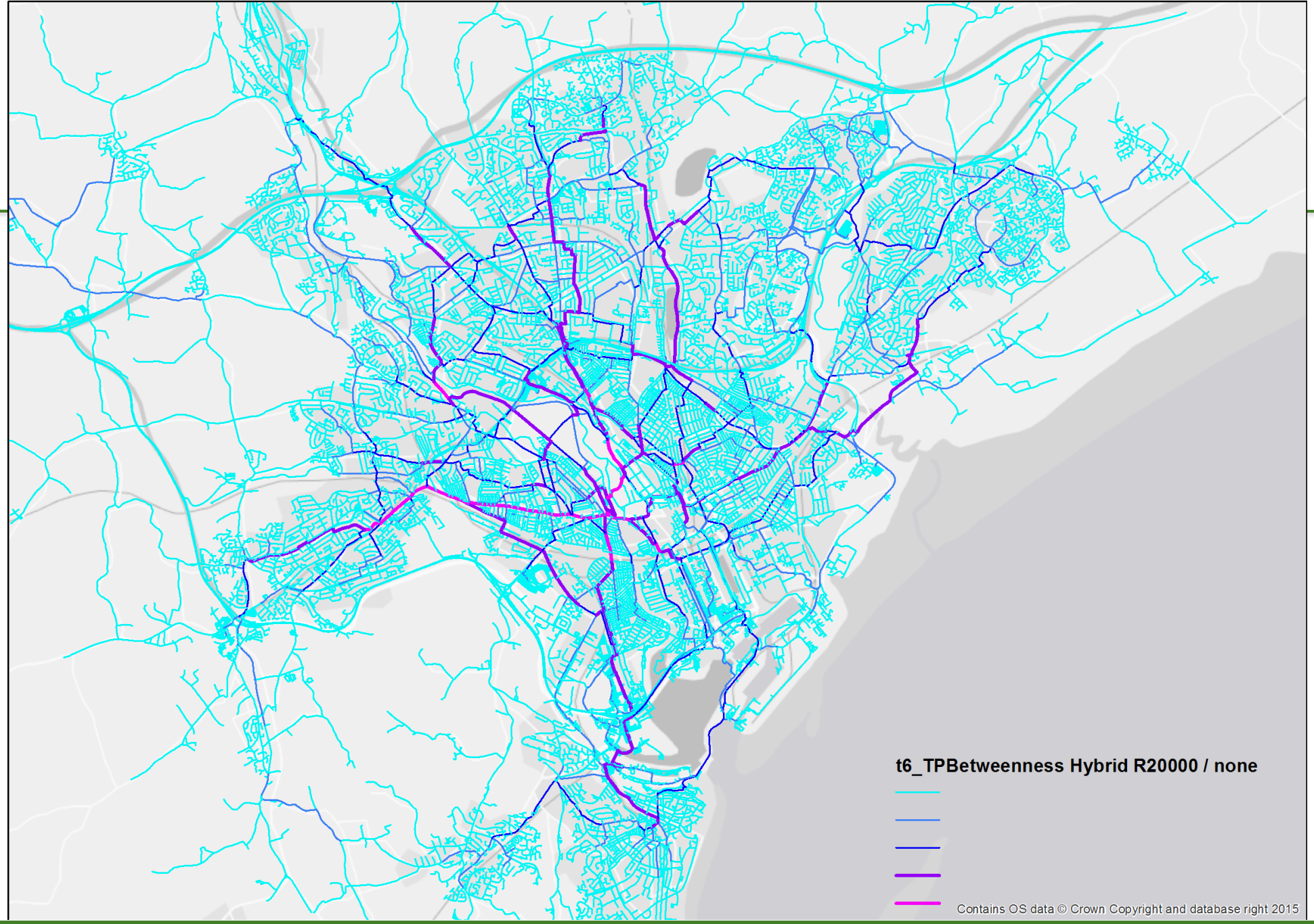

t4taff_TPBetweenness Hybrid Wp R11000 / non
Contains OS data © Crown Copyright and database right 2015

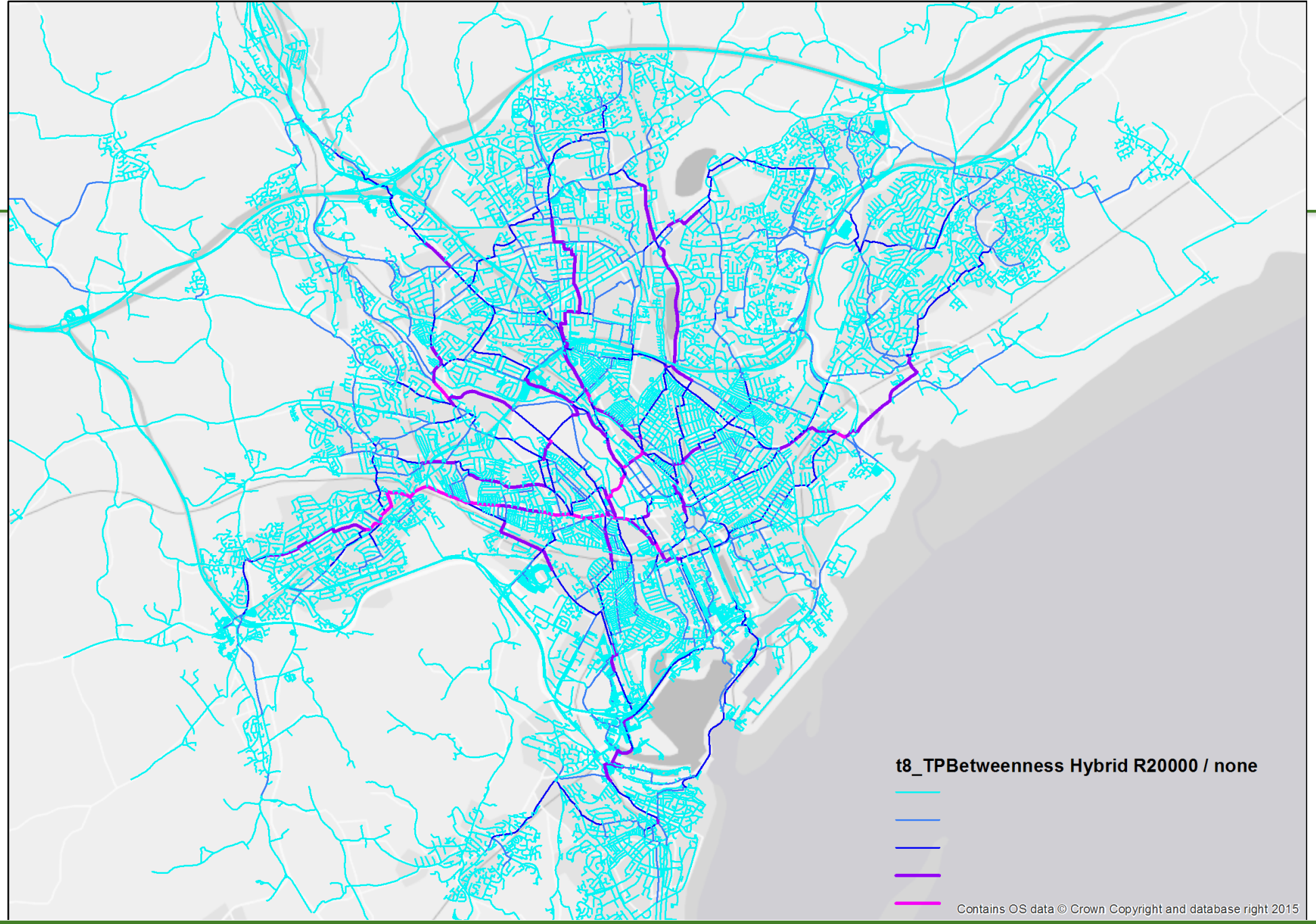

t4taff_TPBetweenness Hybrid Wp R11000 / non
Contains OS data © Crown Copyright and database right 2015


t6taff_TPBetweenness Hybrid Wp R11000 / non
Contains OS data © Crown Copyright and database right 2015


t8taff_TPBetweenness Hybrid Wp R11000 / non
Contains OS data © Crown Copyright and database right 2015

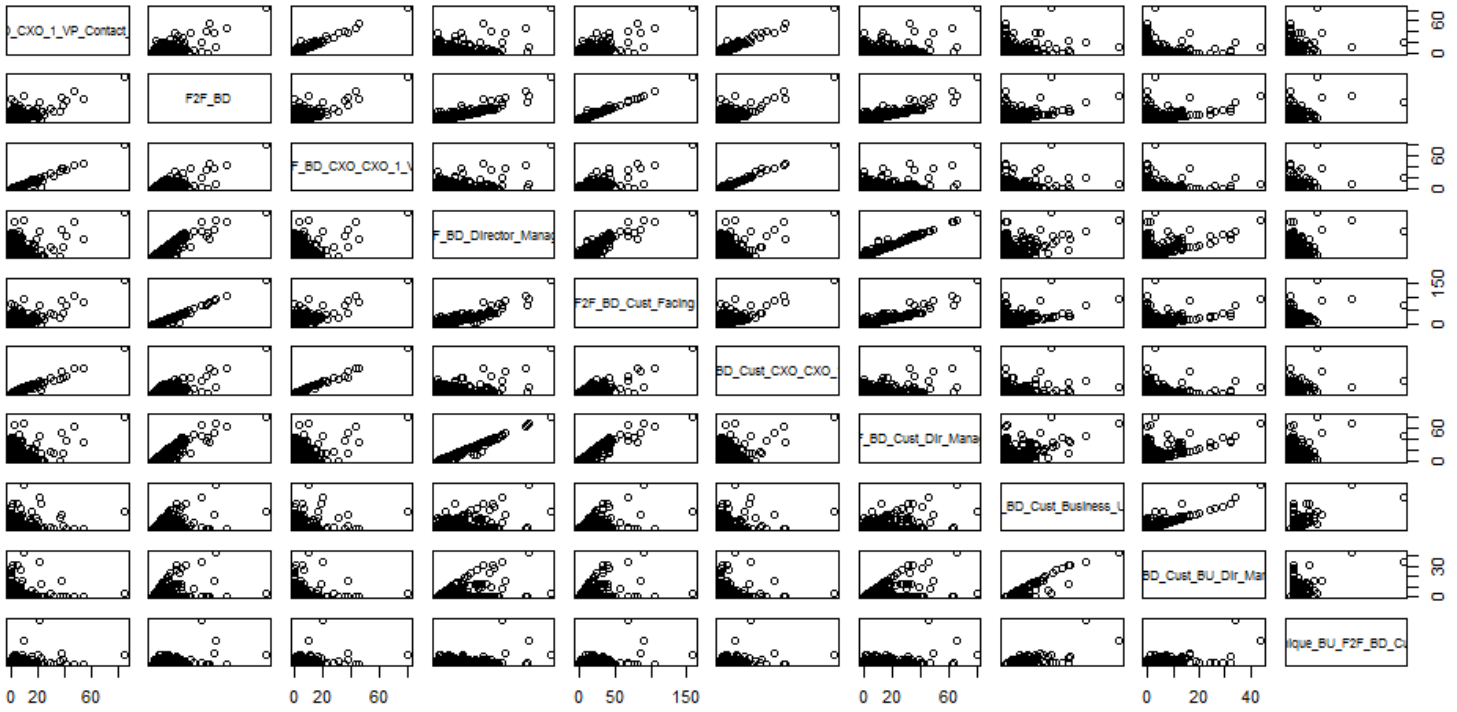

t4_TPBetweenness Hybrid R20000 / none
Contains OS data © Crown Copyright and database right 2015

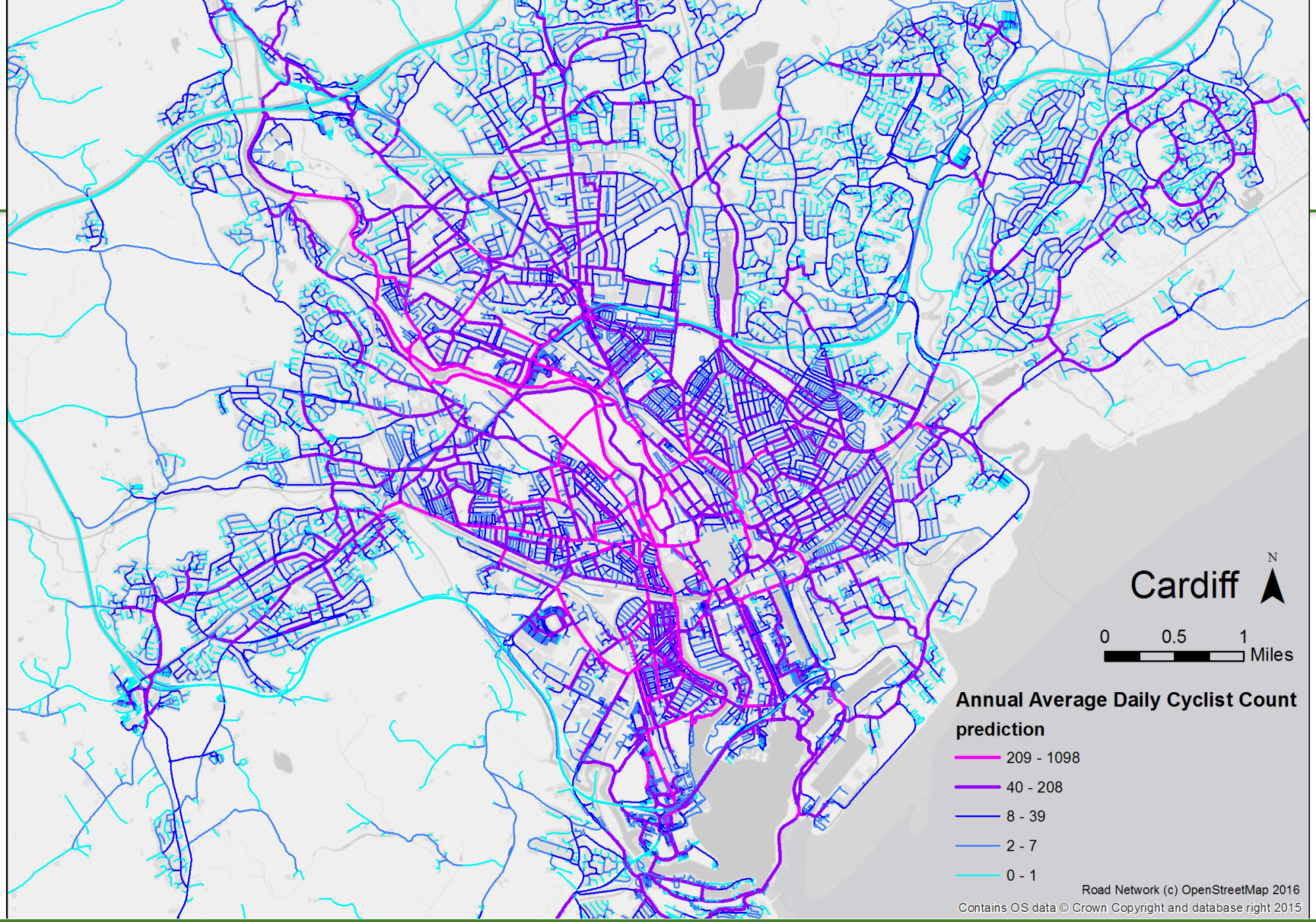

t6_TPBetweenness Hybrid R20000 / none
Contains OS data © Crown Copyright and database right 2015

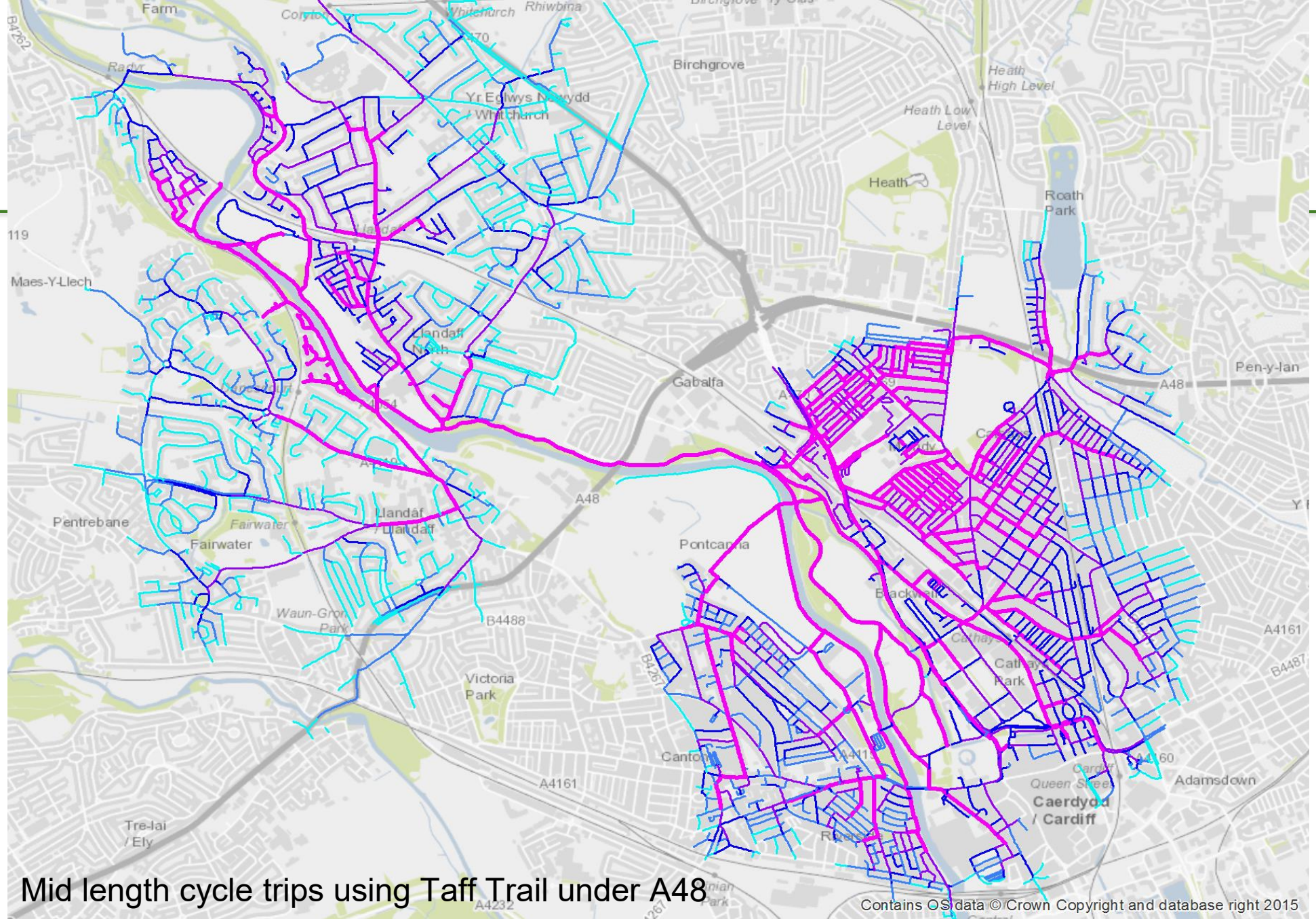

t8_TPBetweenness Hybrid R20000 / none
Contains OS data © Crown Copyright and database right 2015

Predicting flows

- Each betweenness calculation can be seen as a potential cyclist behaviour / independent variable
- Fit linear model to flow data to see what combination of behaviours best explains observations
- High collinearity between variables leads to overfitting on noise generated by differences between similar variables
- Use regularized regression
  - OLS minimizes $\sum(y - \beta X)^2$
  - ridge minimizes $\sum(y - \beta X)^2 + \lambda\sum\beta^2$
  - lasso minimizes $\sum(y - \beta X)^2 + \lambda\sum|\beta|$
- Cross-validation to select optimal λ and prevent overfit
- R^2 (weighted to reduce effect of large flows) = 0.78

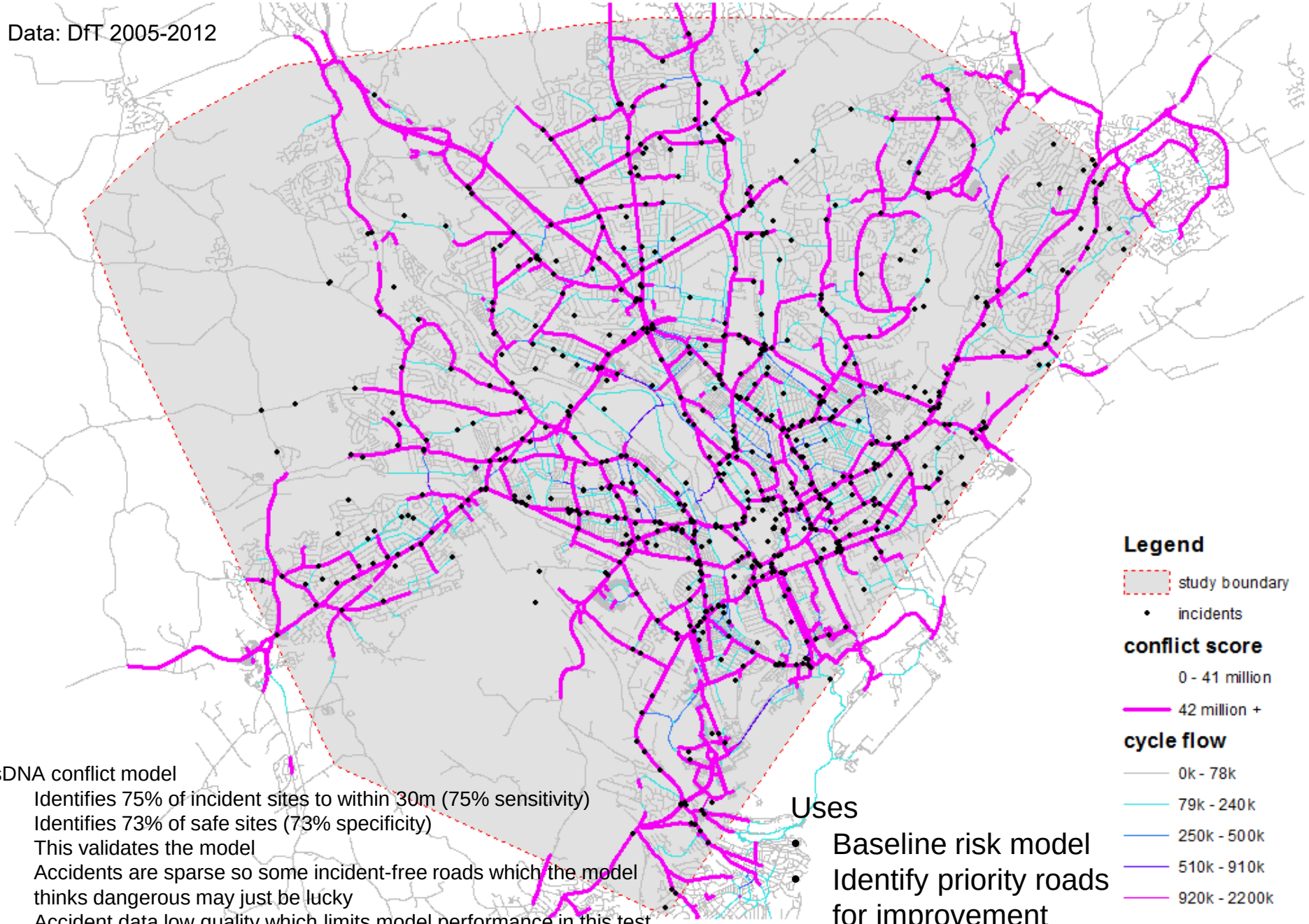



Cardiff
N
0
0.5
1
Miles
Annual Average Daily Cyclist Count
prediction
209 - 1098
40 - 208
8 - 39
2 - 7
0 - 1
Road Network (c) OpenStreetMap 2016
Contains OS data © Crown Copyright and database right 2015



Mid length cycle trips using Taff Trail under A48

Data: DfT 2005-2012

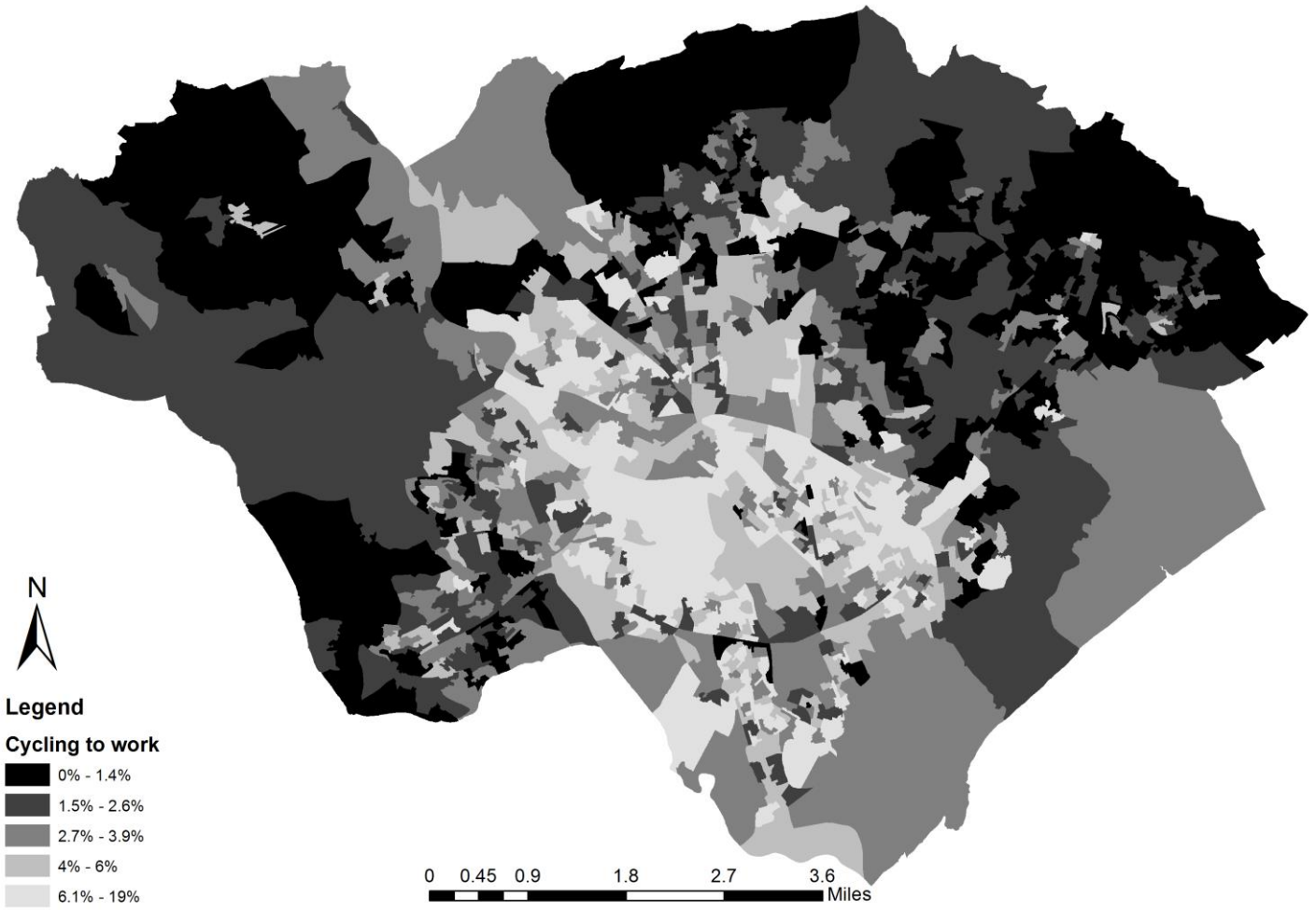


sDNA conflict model

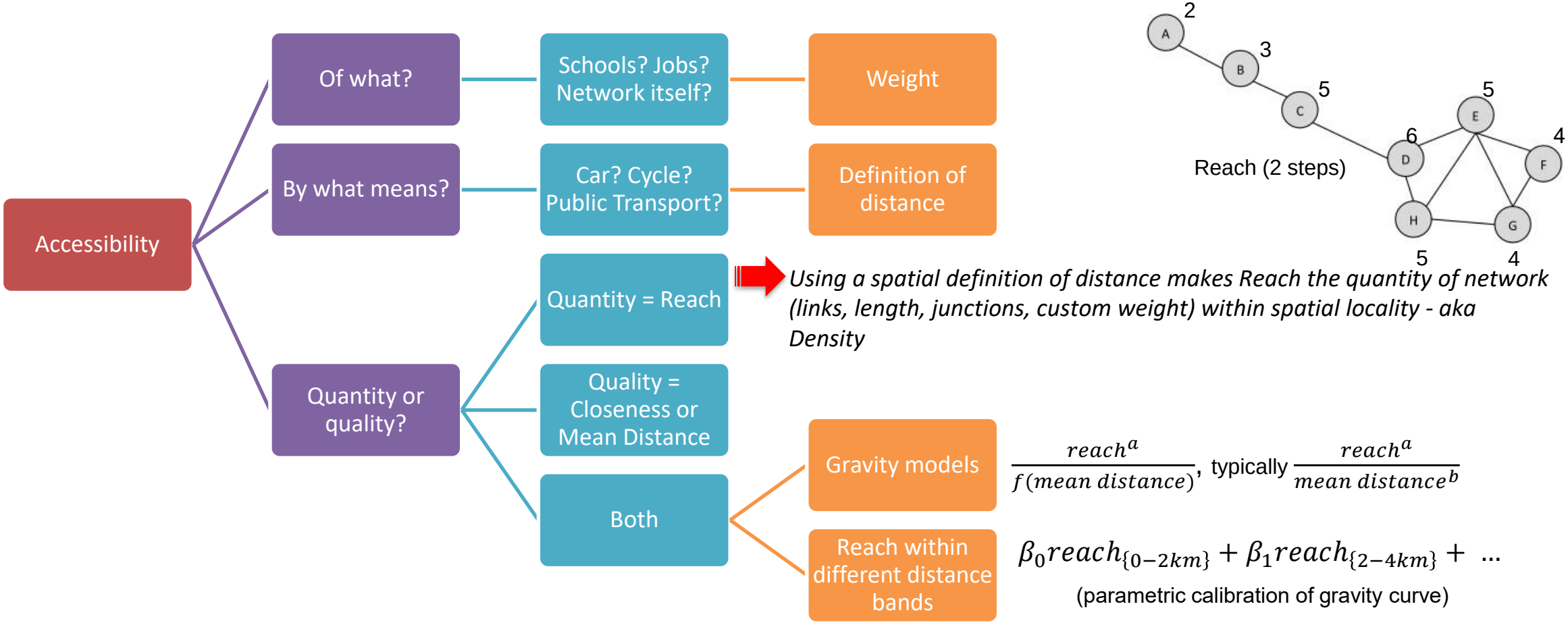
- Identifies 75% of incident sites to within 30m (75% sensitivity)
- Identifies 73% of safe sites (73% specificity)
- This validates the model
- Accidents are sparse so some incident-free roads which the model thinks dangerous may just be lucky
- Accident data low quality which limits model performance in this test

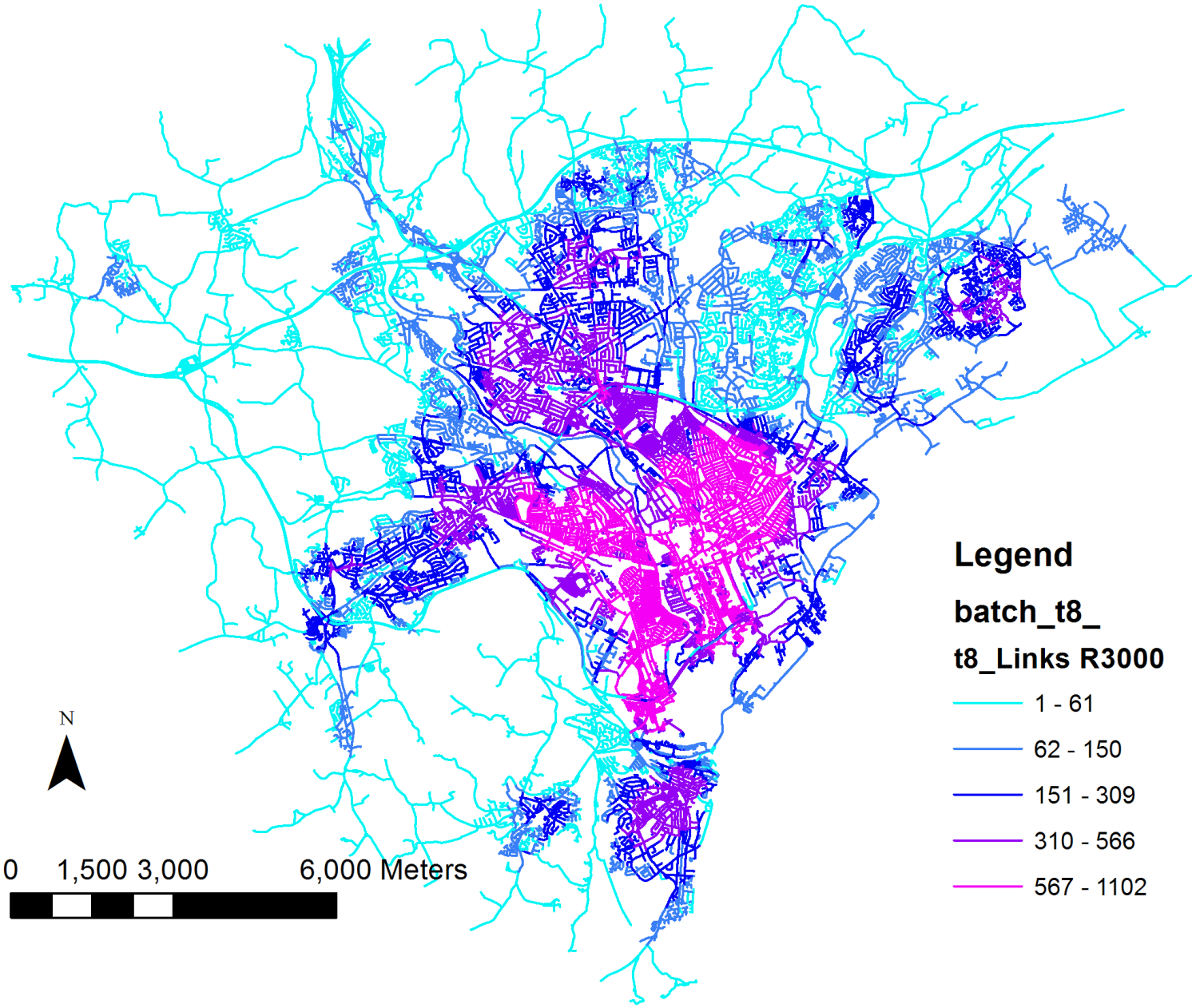
Uses

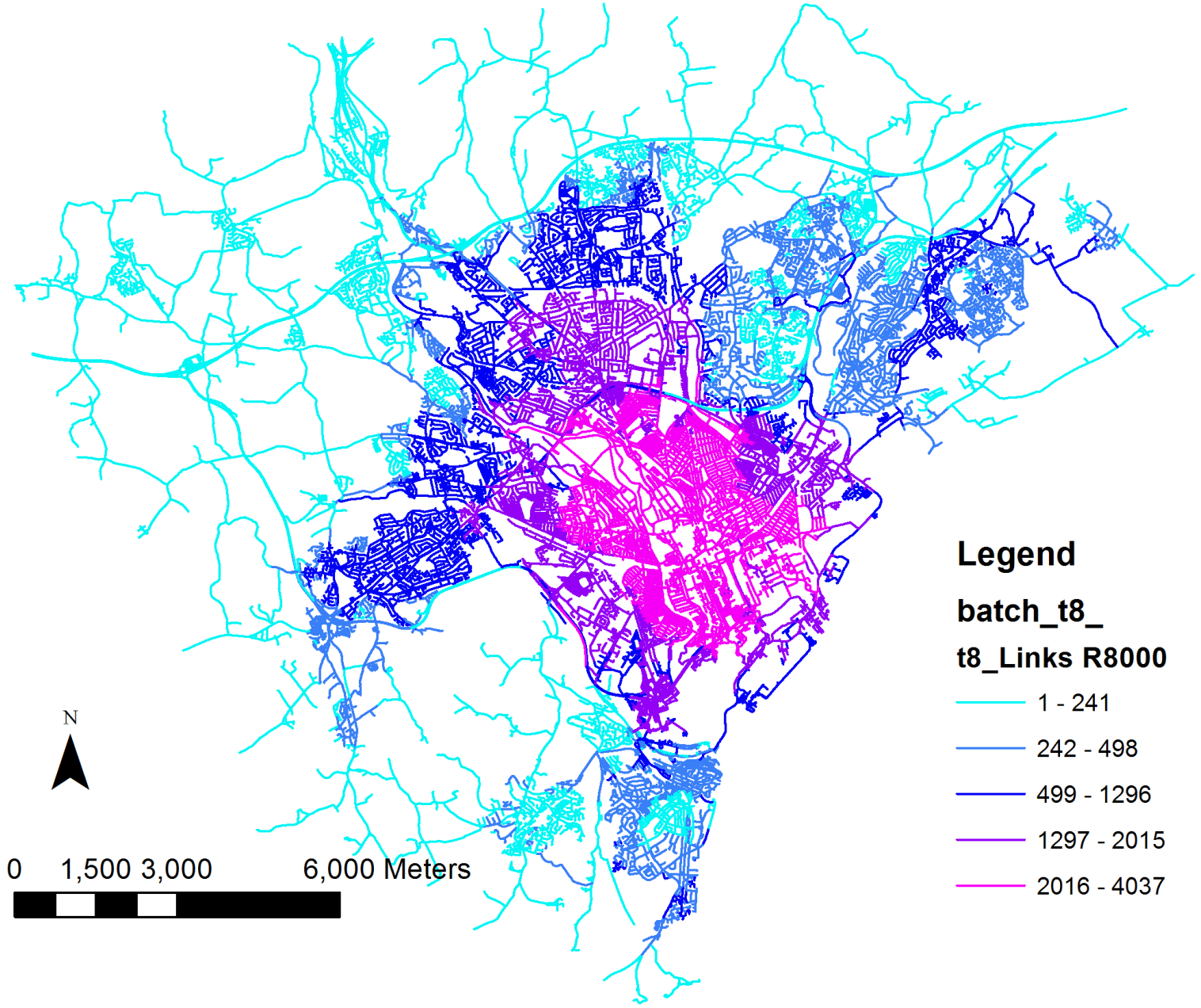
- Baseline risk model
- Identify priority roads for improvement

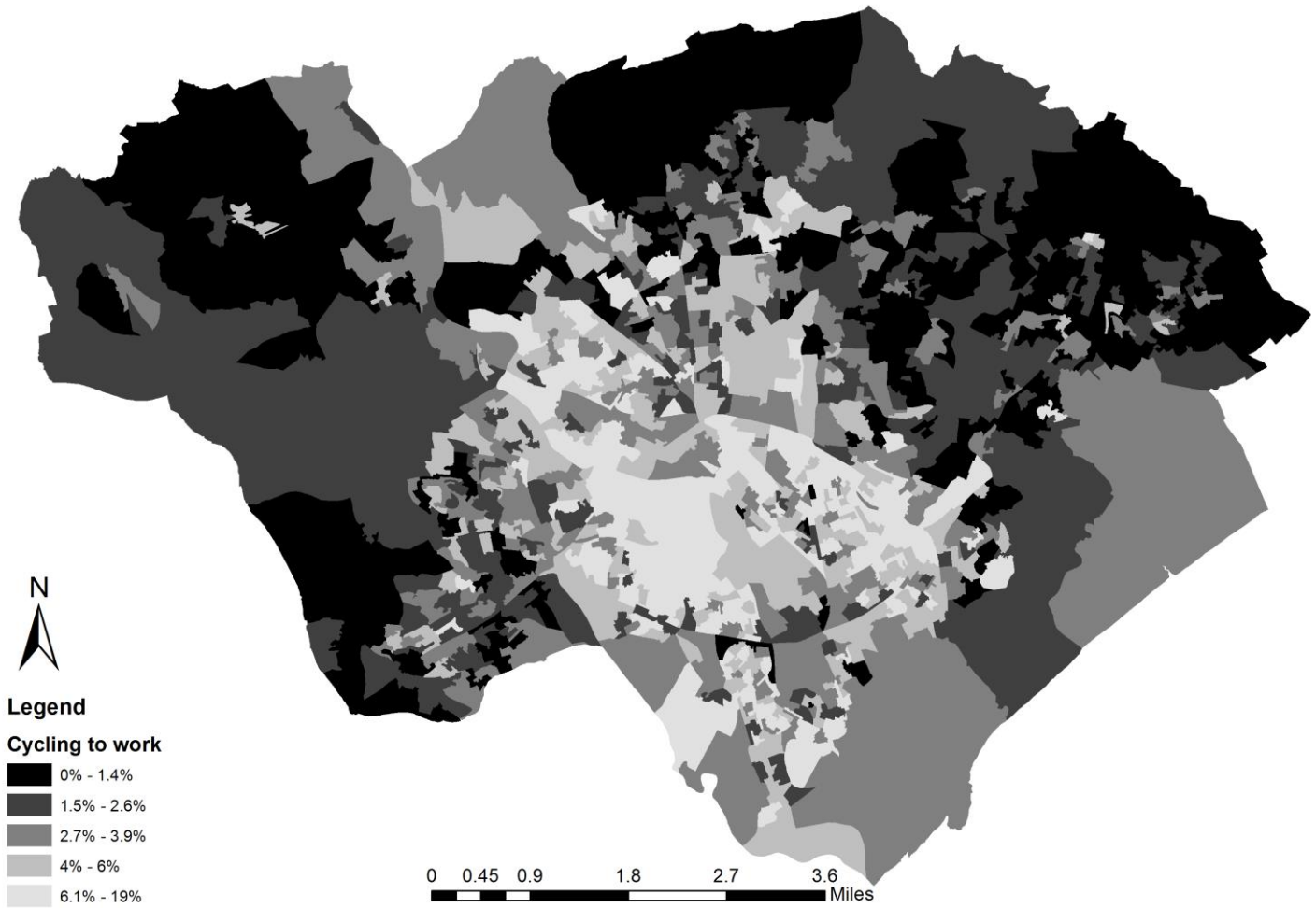

N
Legend
Cycling to work
0% - 1.4%
1.5% - 2.6%
2.7% - 3.9%
4% - 6%
6.1% - 19%
0 0.45 0.9 1.8 2.7 3.6
Miles

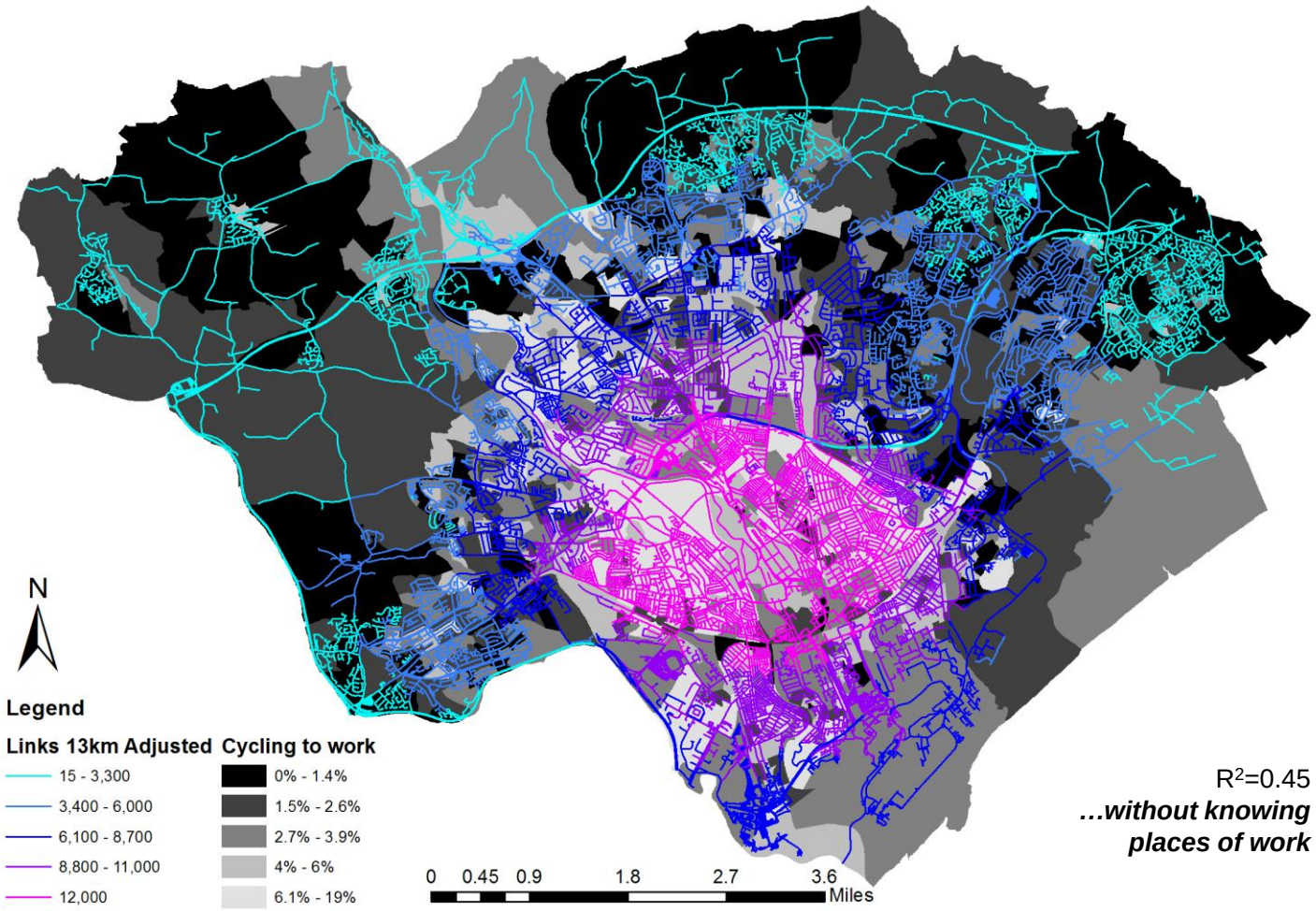
Measuring Accessibility




N
0 1,500 3,000 6,000 Meters
Legend
batch_t8_
t8_Links R3000
1 - 61
62 - 150
151 - 309
310 - 566
567 - 1102


N
0 1,500 3,000 6,000 Meters
Legend
batch_t8_
t8_Links R8000
1 - 241
242 - 498
499 - 1296
1297 - 2015
2016 - 4037

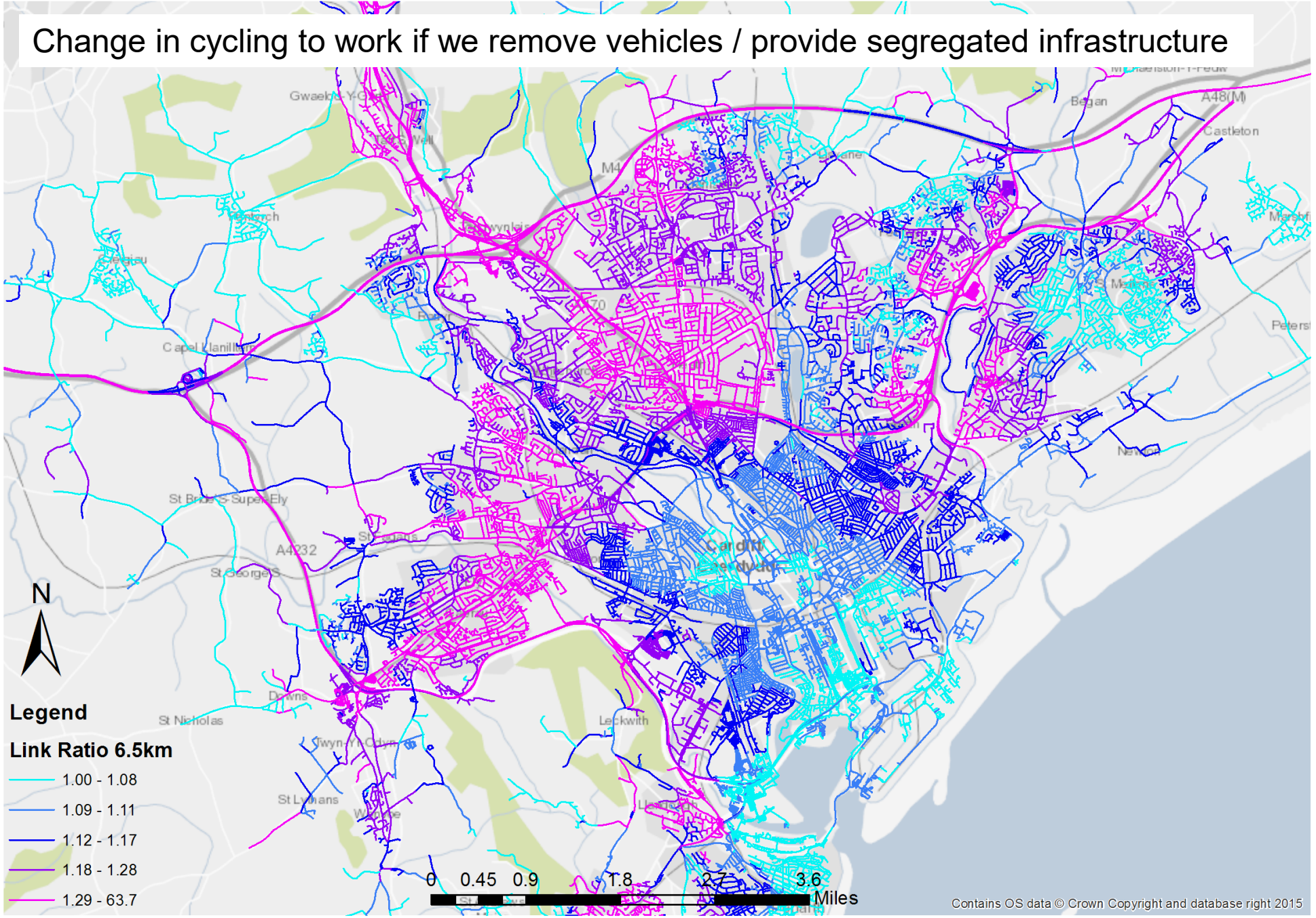

N
Legend
Cycling to work
0% - 1.4%
1.5% - 2.6%
2.7% - 3.9%
4% - 6%
6.1% - 19%
0 0.45 0.9 1.8 2.7 3.6
Miles



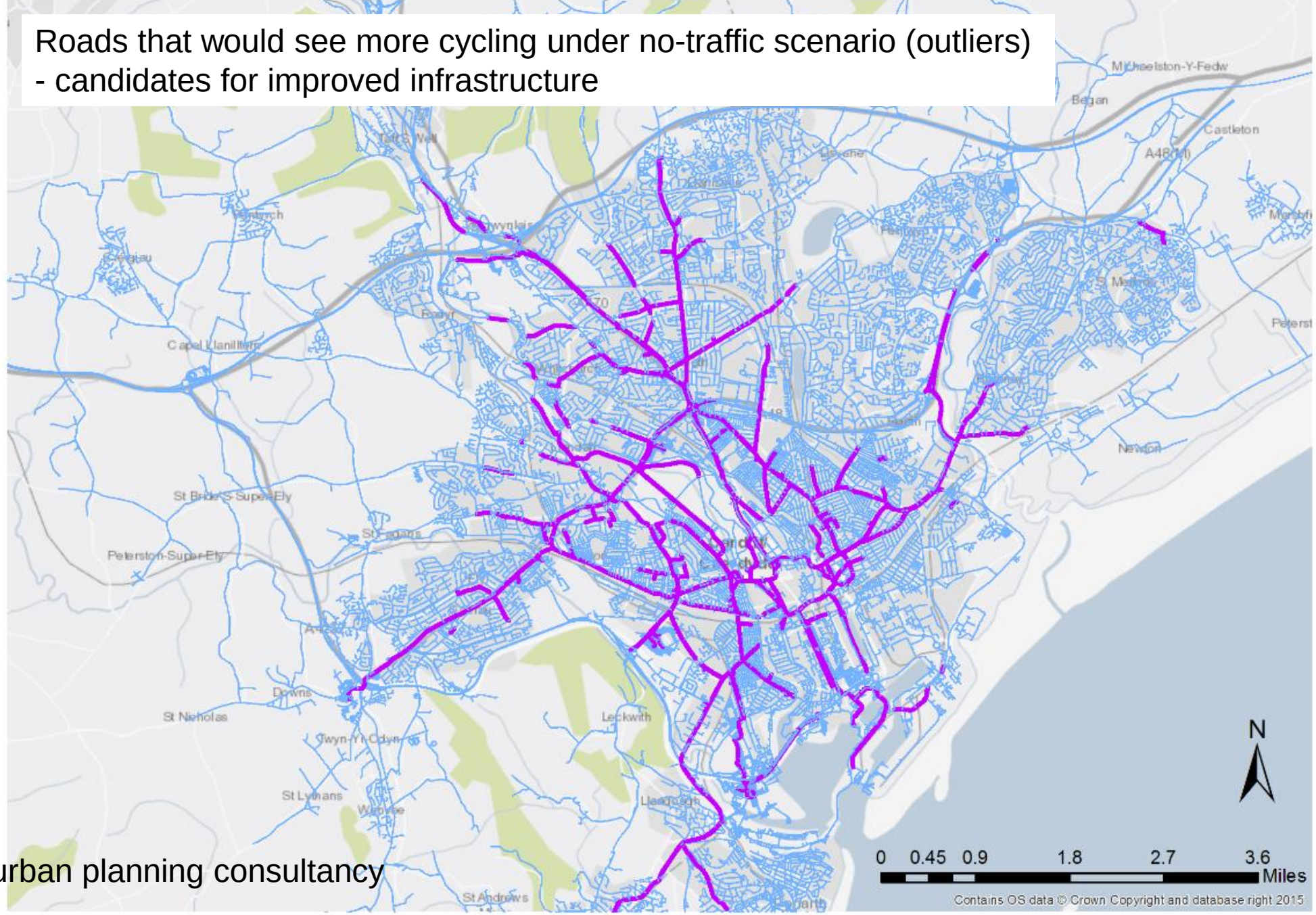
R^2=0.45

…without knowing places of work

Change in cycling to work if we remove vehicles / provide segregated infrastructure



Roads that would see more cycling under no-traffic scenario (outliers)
- candidates for improved infrastructure



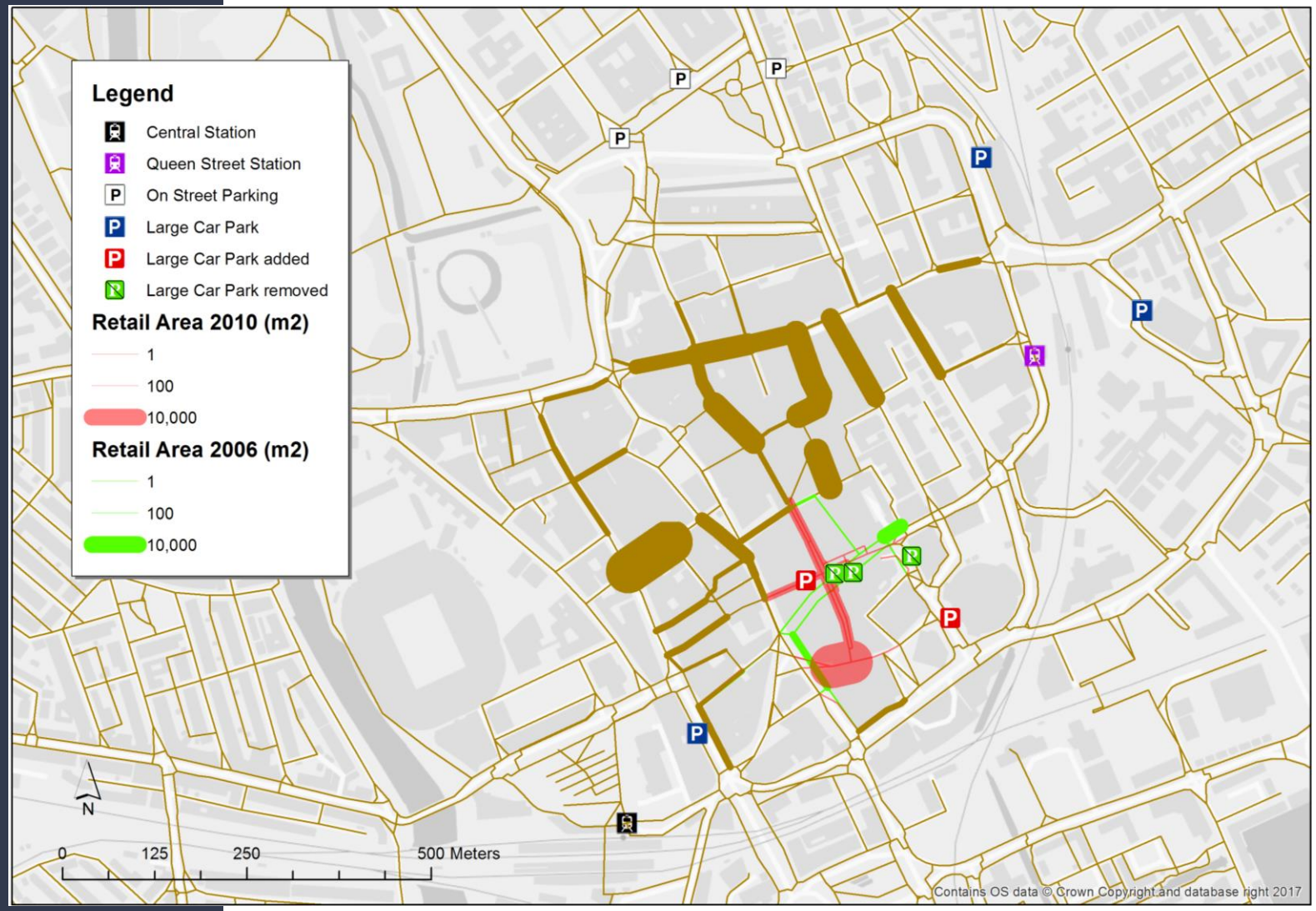
Uses: Arup urban planning consultancy

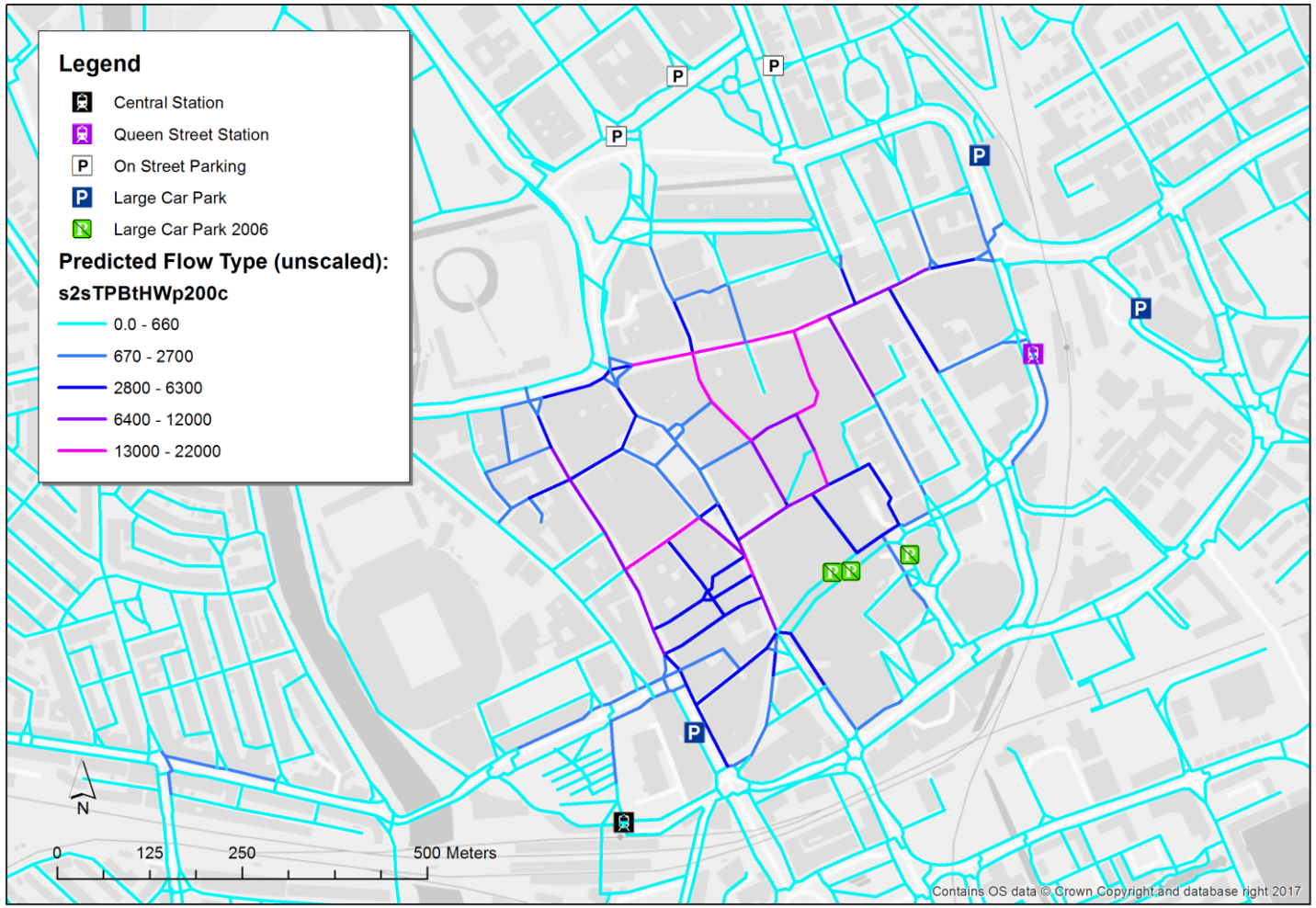
Cardiff model of walking

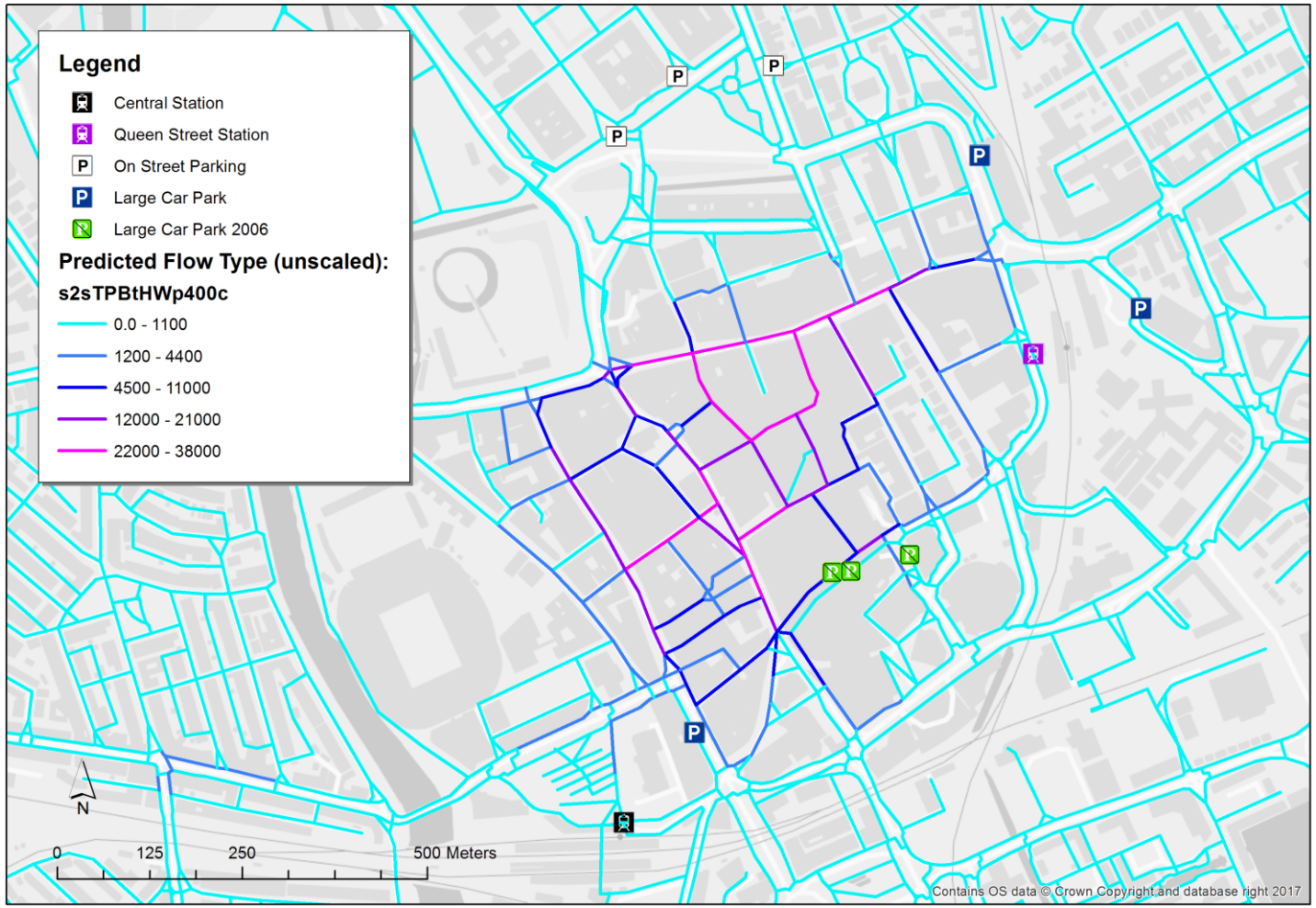
Longitudinal test against Cardiff data

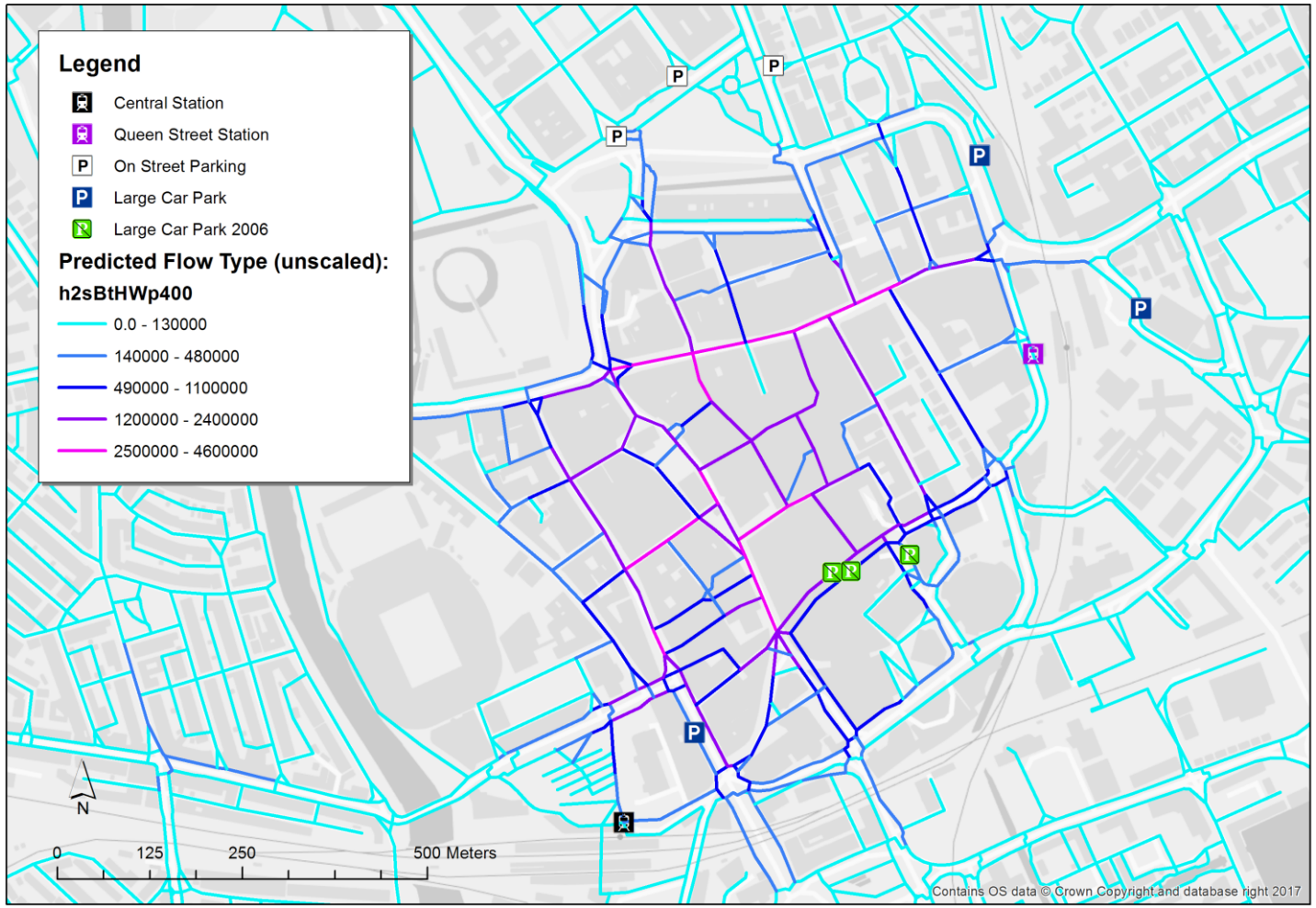
Major city centre redevelopment
2007-2010

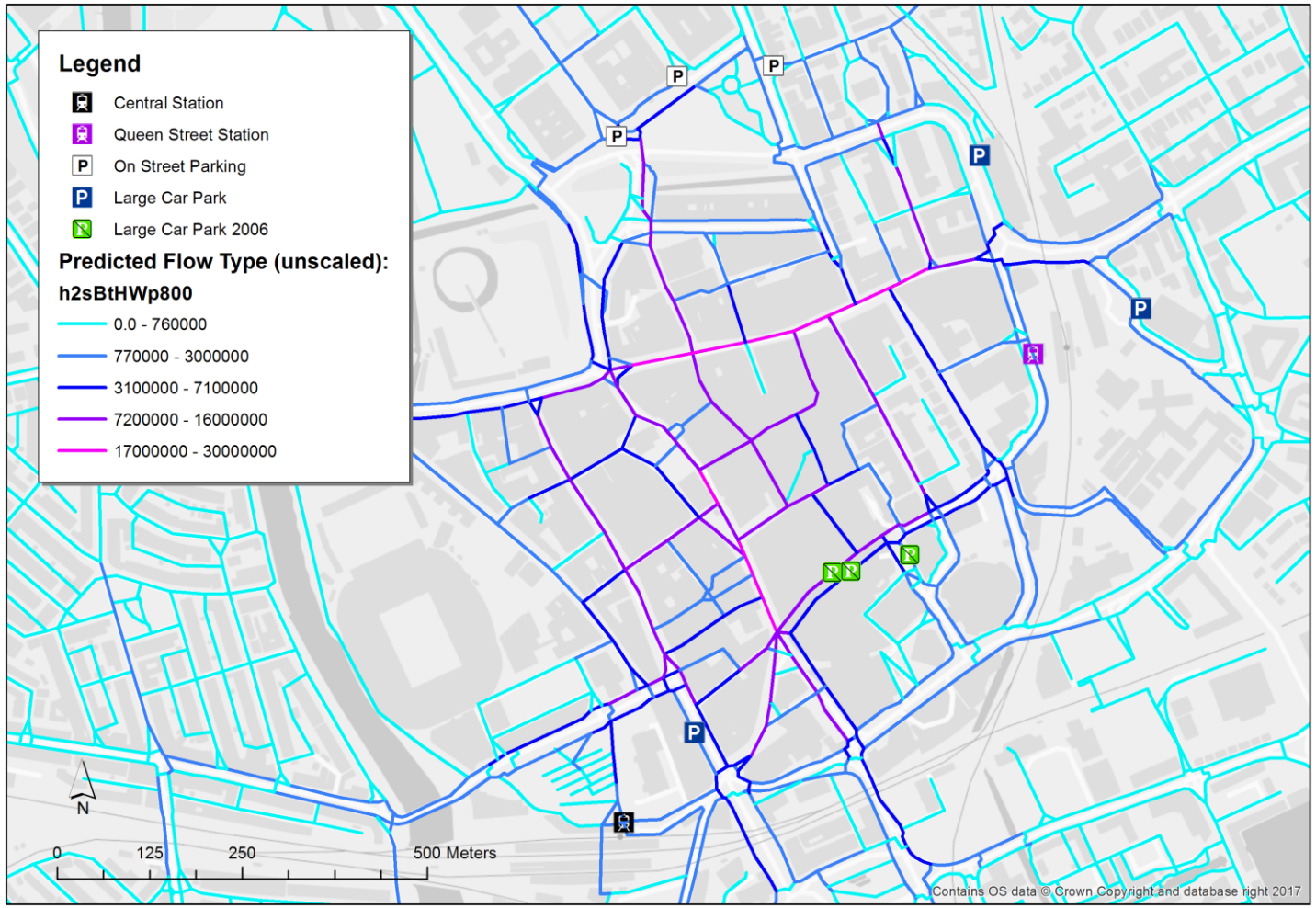
Data Innovation Institute seedcorn project with
Ian Harvey
Scott Orford
Alain Chiaradia

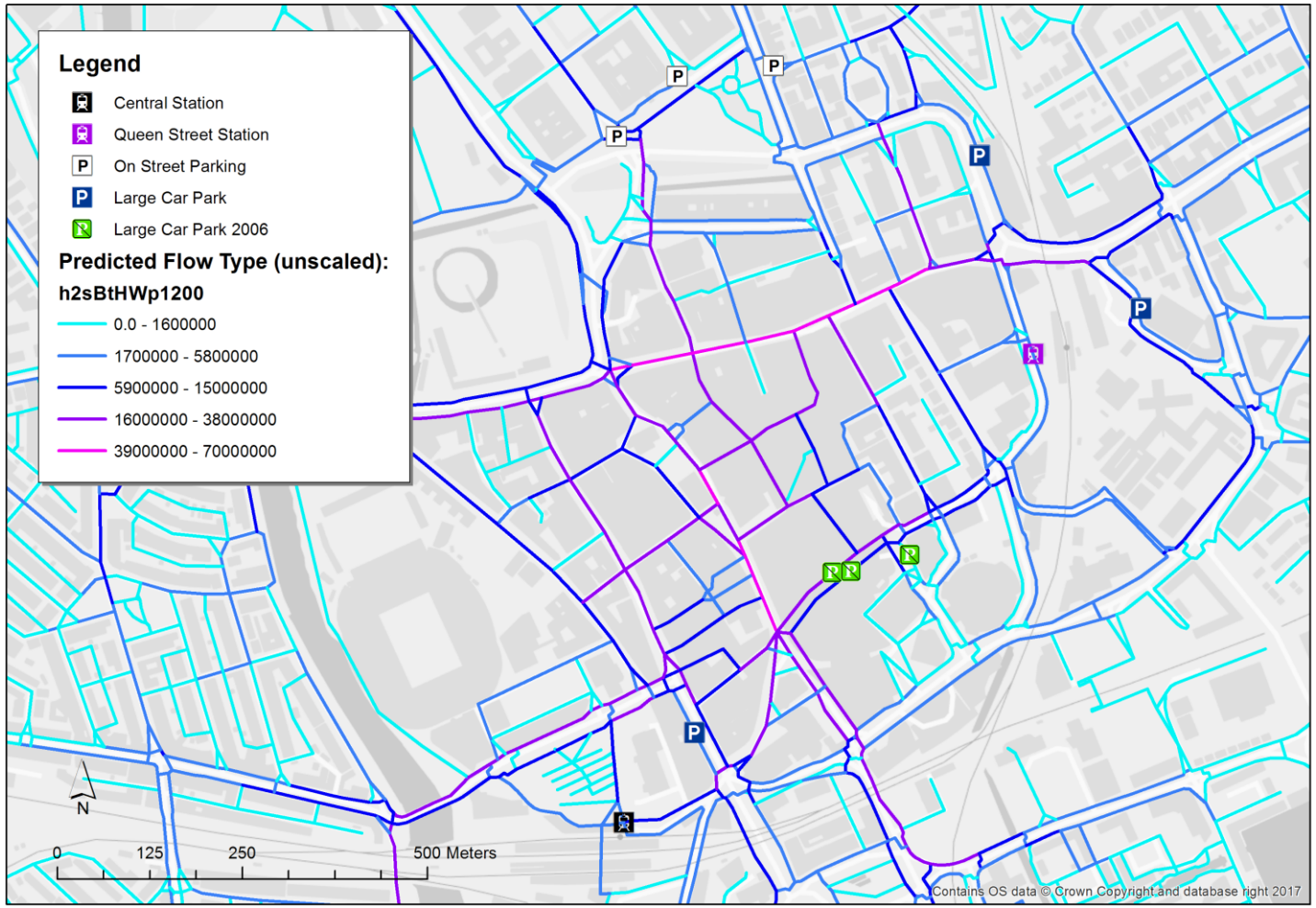


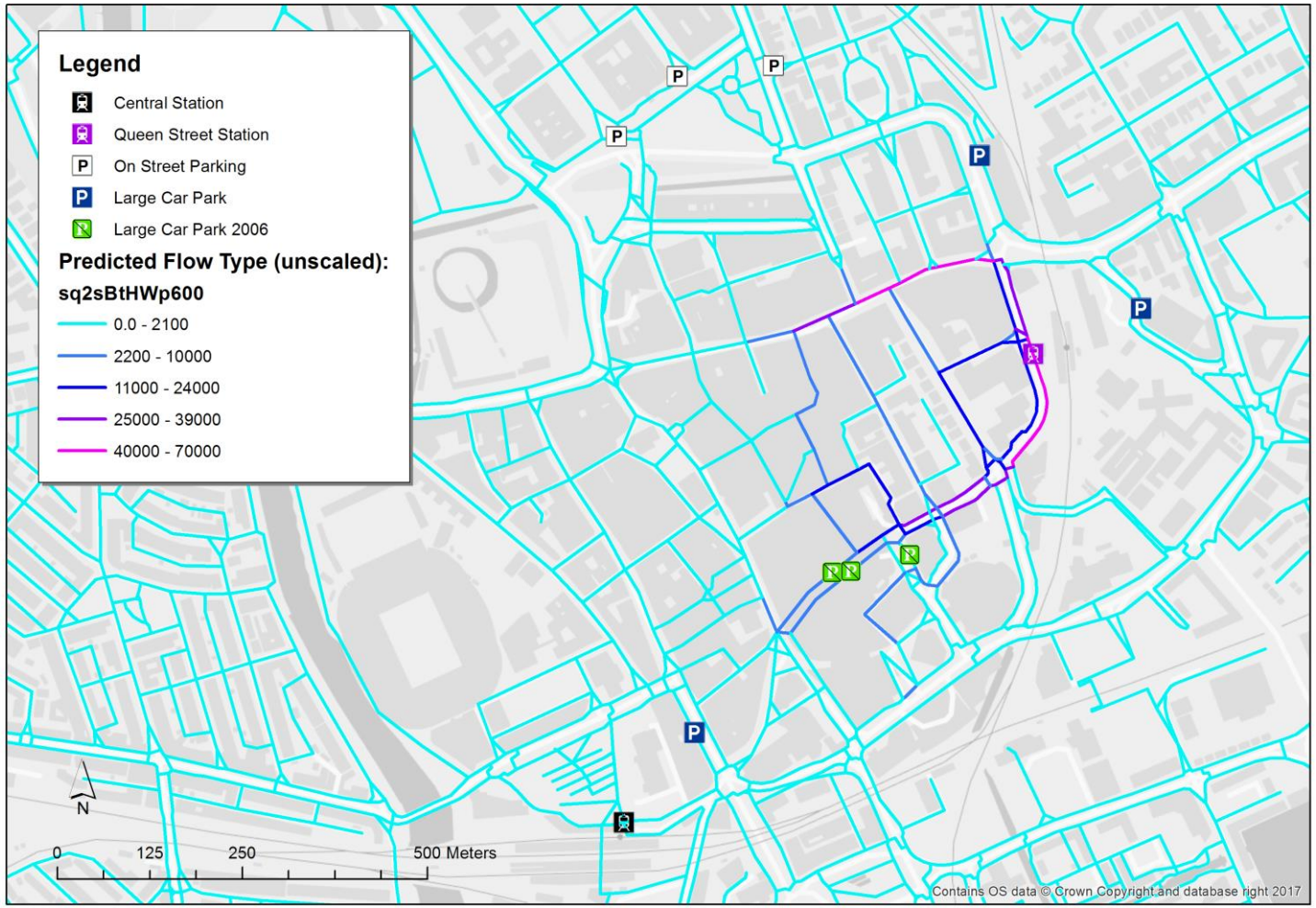

Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
s2sTPBtHWp200c
0.0 - 660
670 - 2700
2800 - 6300
6400 - 12000
13000 - 22000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017

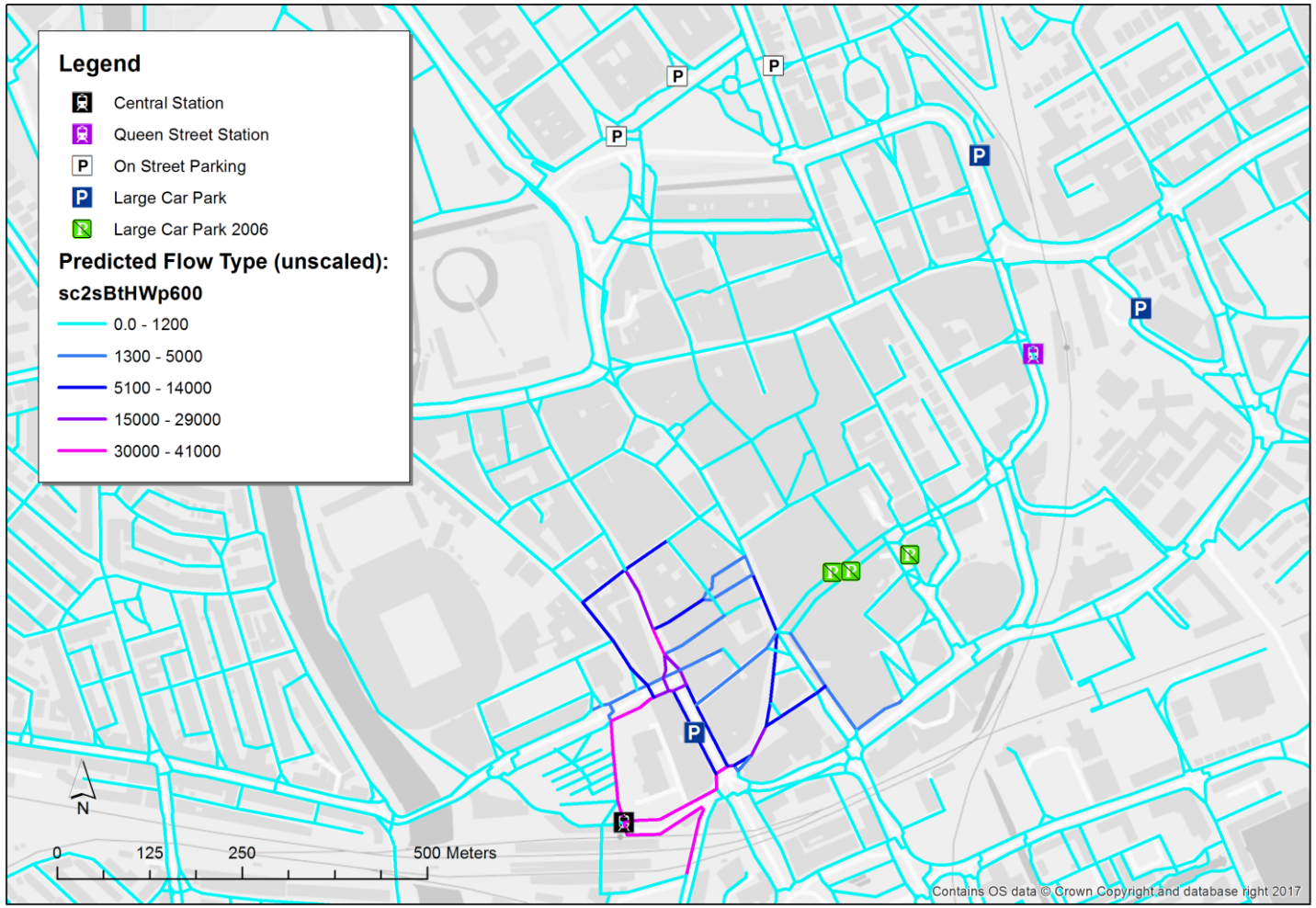

Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
s2sTPBtHWp400c
0.0 - 1100
1200 - 4400
4500 - 11000
12000 - 21000
22000 - 38000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017

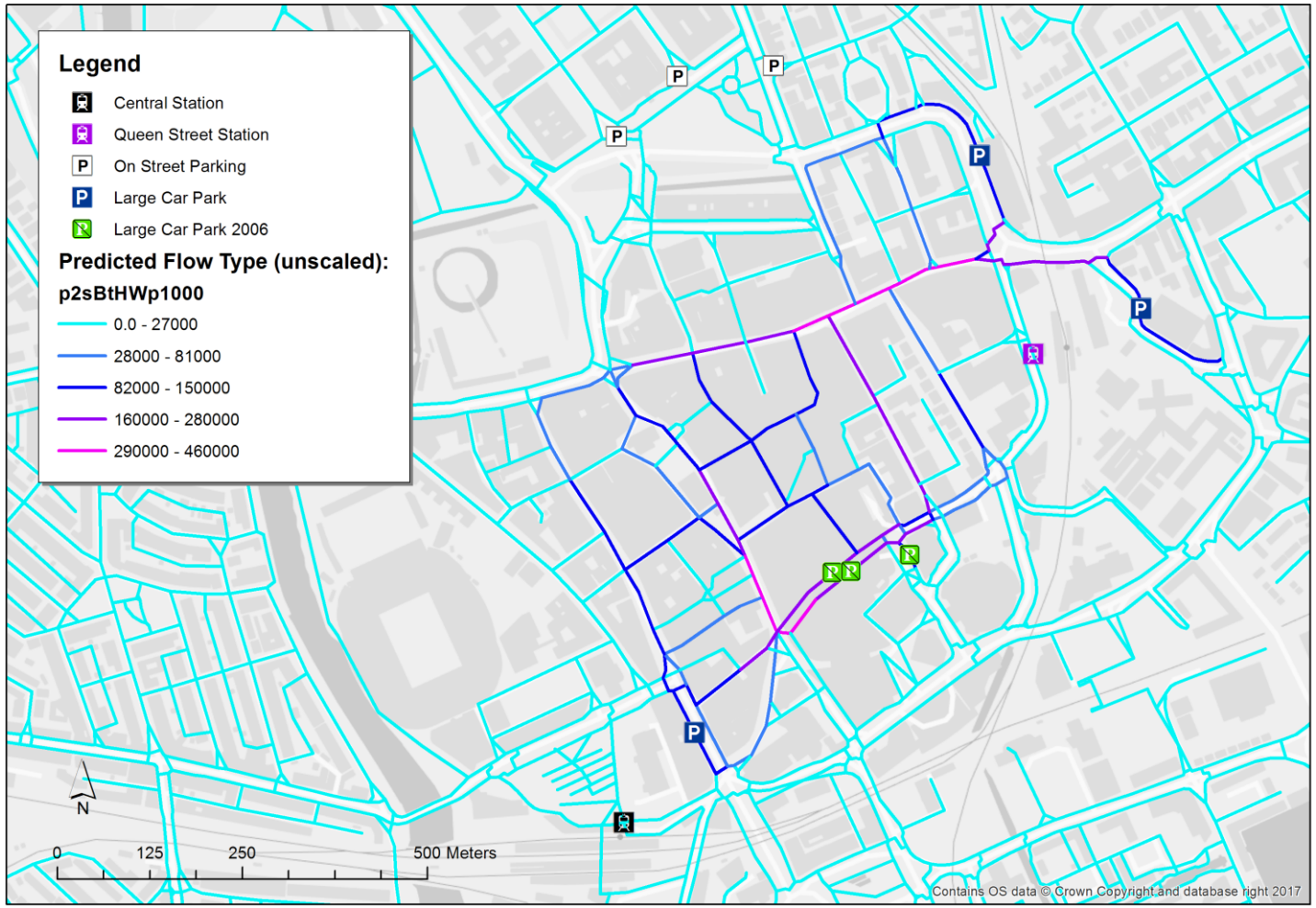

Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
h2sBtHWp400
0.0 - 130000
140000 - 480000
490000 - 1100000
1200000 - 2400000
2500000 - 4600000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017

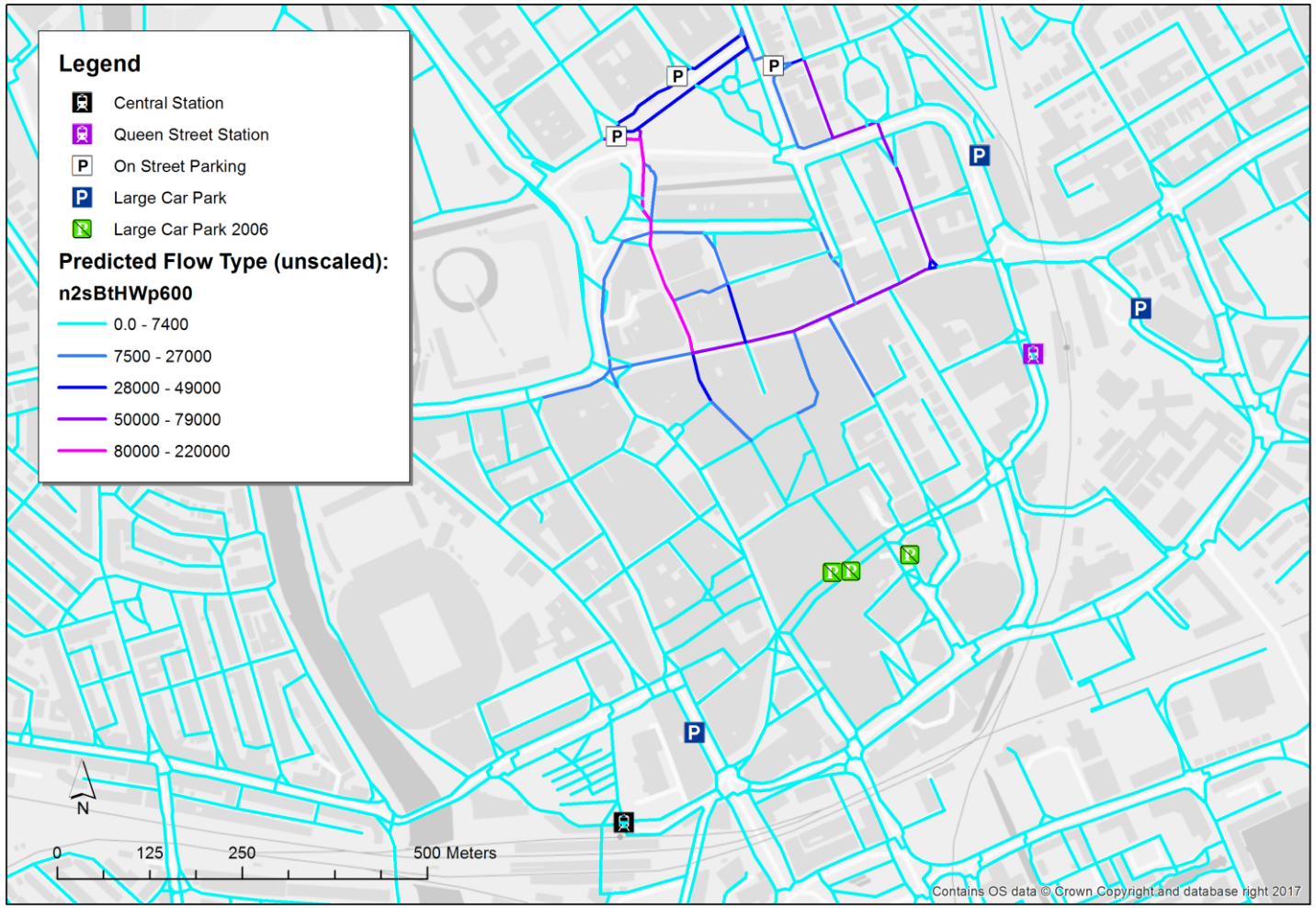

Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
h2sBtHWp800
0.0 - 760000
770000 - 3000000
3100000 - 7100000
7200000 - 16000000
17000000 - 30000000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017

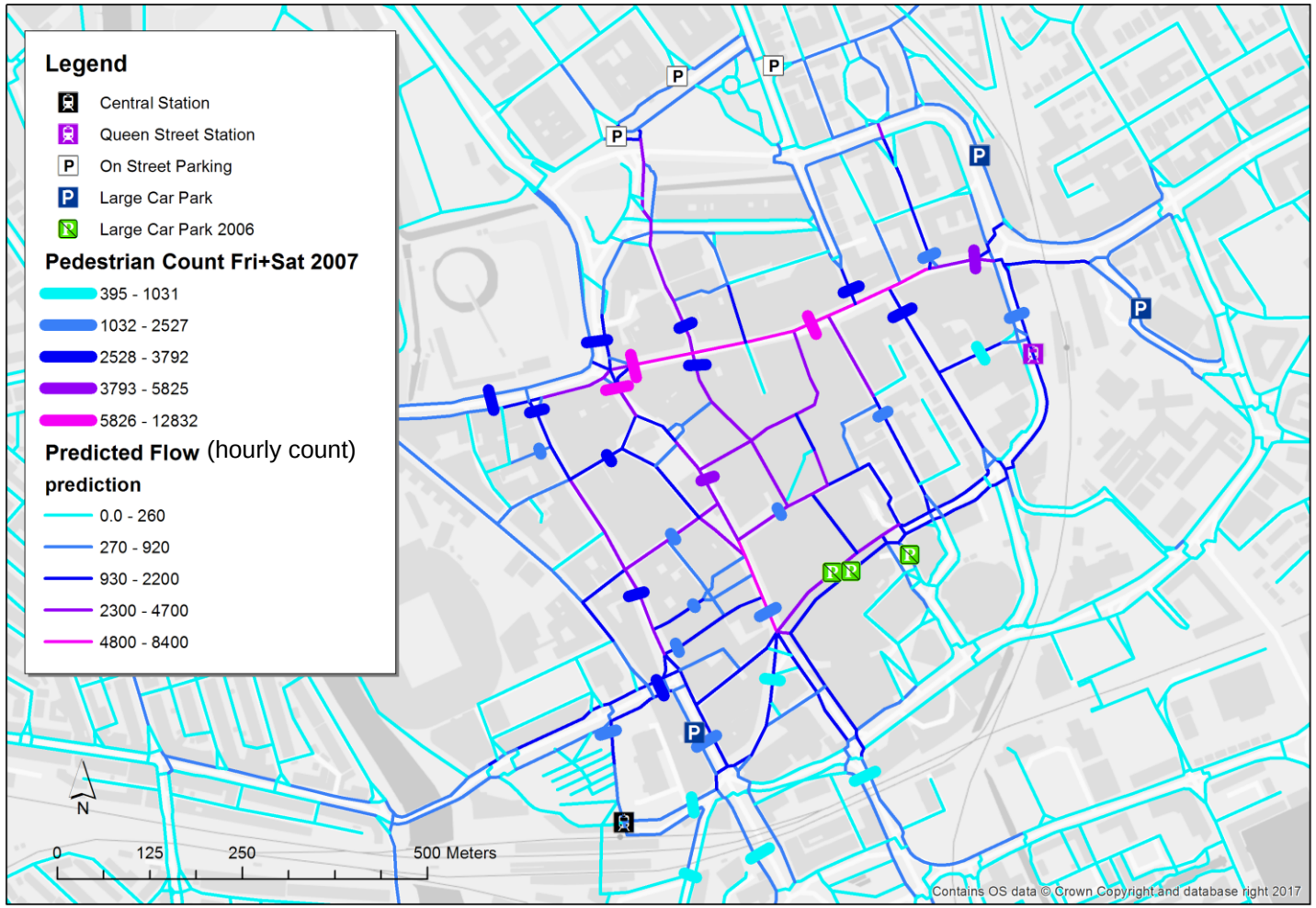

Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
h2sBtHWp1200
0.0 - 1600000
1700000 - 5800000
5900000 - 15000000
16000000 - 38000000
39000000 - 70000000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017


Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
sq2sBtHWp600
0.0 - 2100
2200 - 10000
11000 - 24000
25000 - 39000
40000 - 70000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017


Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
sc2sBtHWp600
0.0 - 1200
1300 - 5000
5100 - 14000
15000 - 29000
30000 - 41000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017

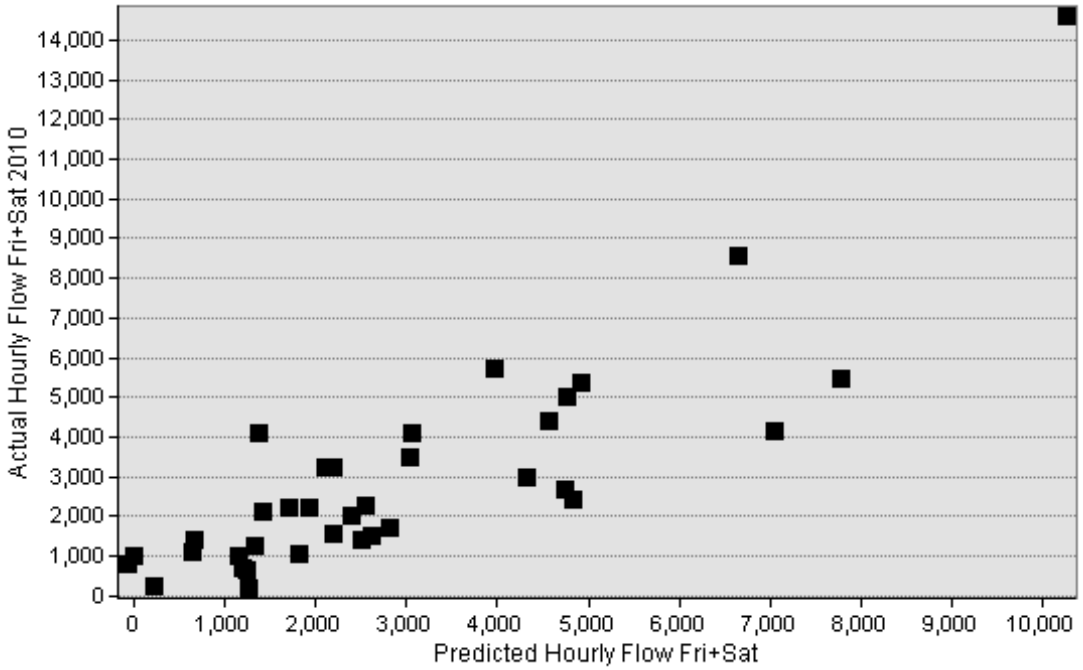

Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
p2sBtHWp1000
0.0 - 27000
28000 - 81000
82000 - 150000
160000 - 280000
290000 - 460000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017


Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Predicted Flow Type (unscaled):
n2sBtHWp600
0.0 - 7400
7500 - 27000
28000 - 49000
50000 - 79000
80000 - 220000
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017


Legend
Central Station
Queen Street Station
On Street Parking
Large Car Park
Large Car Park 2006
Pedestrian Count Fri+Sat 2007
395 - 1031
1032 - 2527
2528 - 3792
3793 - 5825
5826 - 12832
Predicted Flow (hourly count)
prediction
0.0 - 260
270 - 920
930 - 2200
2300 - 4700
4800 - 8400
N
0
125
250
500 Meters
Contains OS data © Crown Copyright and database right 2017

Predictive performance

To our knowledge the first time a pedestrian model has been tested in its ability to predict longitudinal change



Year	Null model r2	Incremental model r2	Direct model r2
2008	0.79	n/a	n/a
2009	0.85	n/a	n/a
2010	0.81	0.84	0.72
2011	0.63	0.73	0.45

Users: Arup, Wedderburn, Alan Baxter, Hong Kong, Shanghai, Transport for London

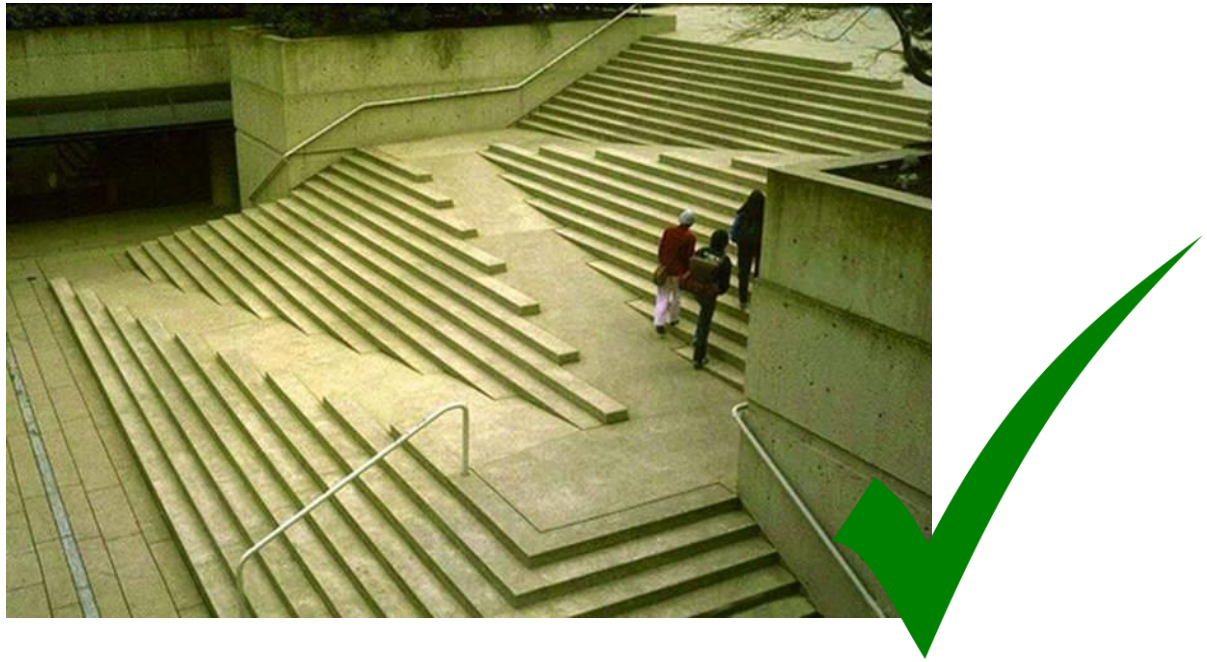
Cardiff City Region: Limited mobilities model

Richard Price MSc Dissertation

No test against data

Mapping severance for people unable to climb steps

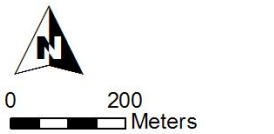
Enabled by audit of infrastructure under Wales Active Travel Act (2014)



Example: Abercynon

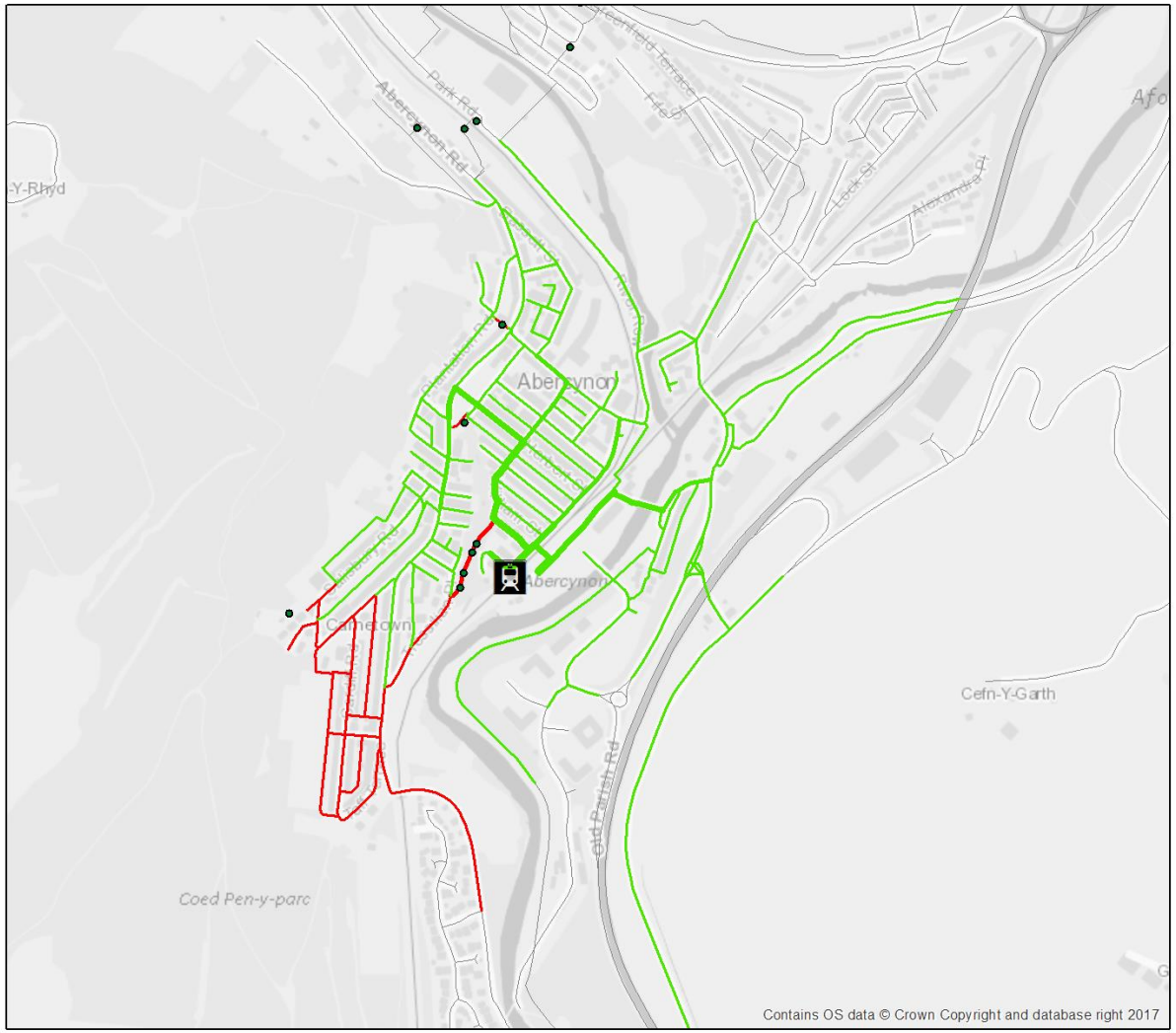
Stepped route provides essential shortcut to south, which cannot otherwise be reached within 800m

Step free route takes detour to north



Legend

- Steps
- Rail Stations
- **800m accessible to all**
- **800m accessible via steps**



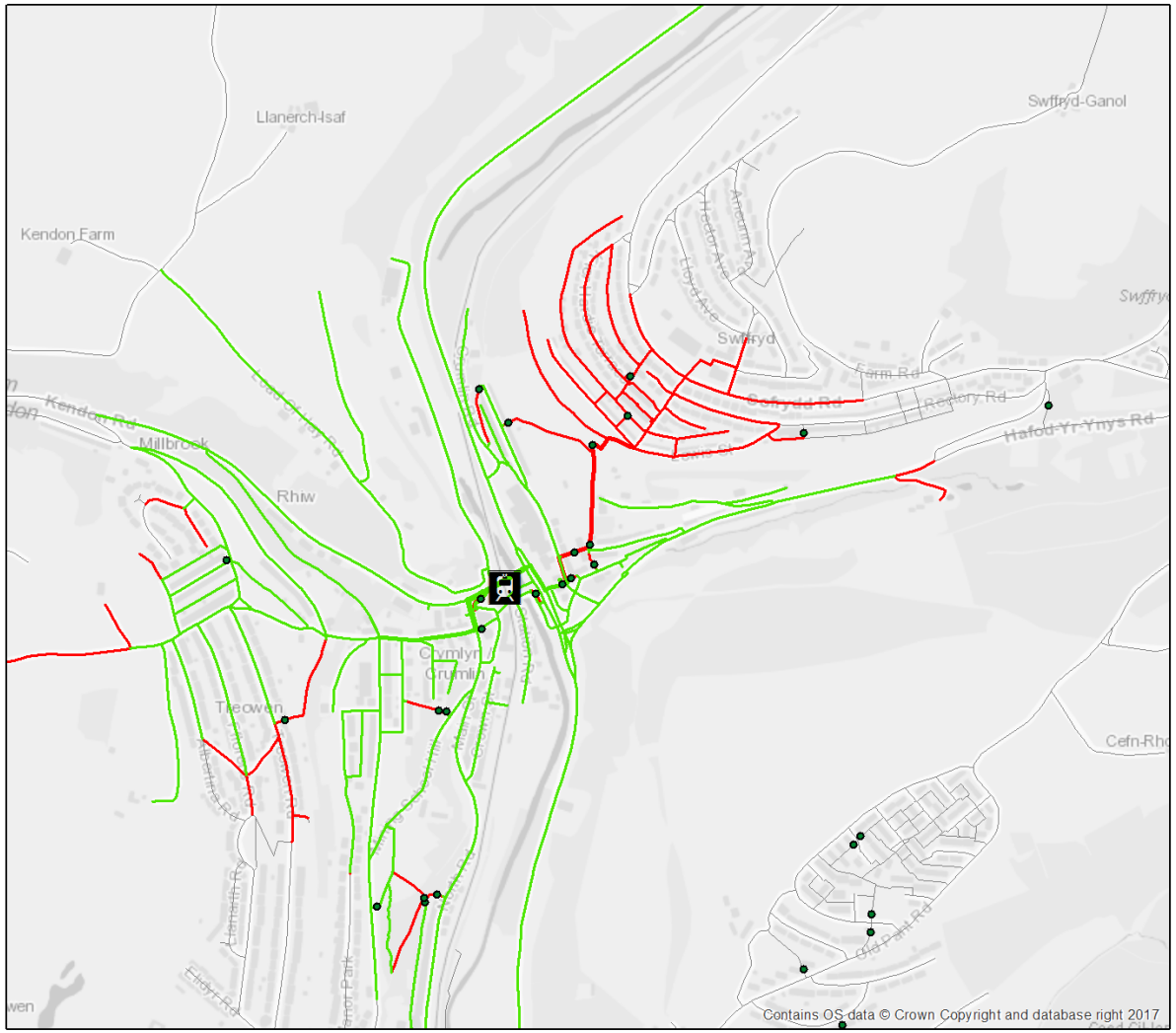
Pilot project – not for planning use

Example: Crymlyn

Stepped route provides essential shortcut to north, which cannot otherwise be reached within 800m

Step free route takes long detour to east

Possibly unavoidable due to topography of valley



0 200 Meters

Legend

- Steps
- Rail Stations

800m accessible to all

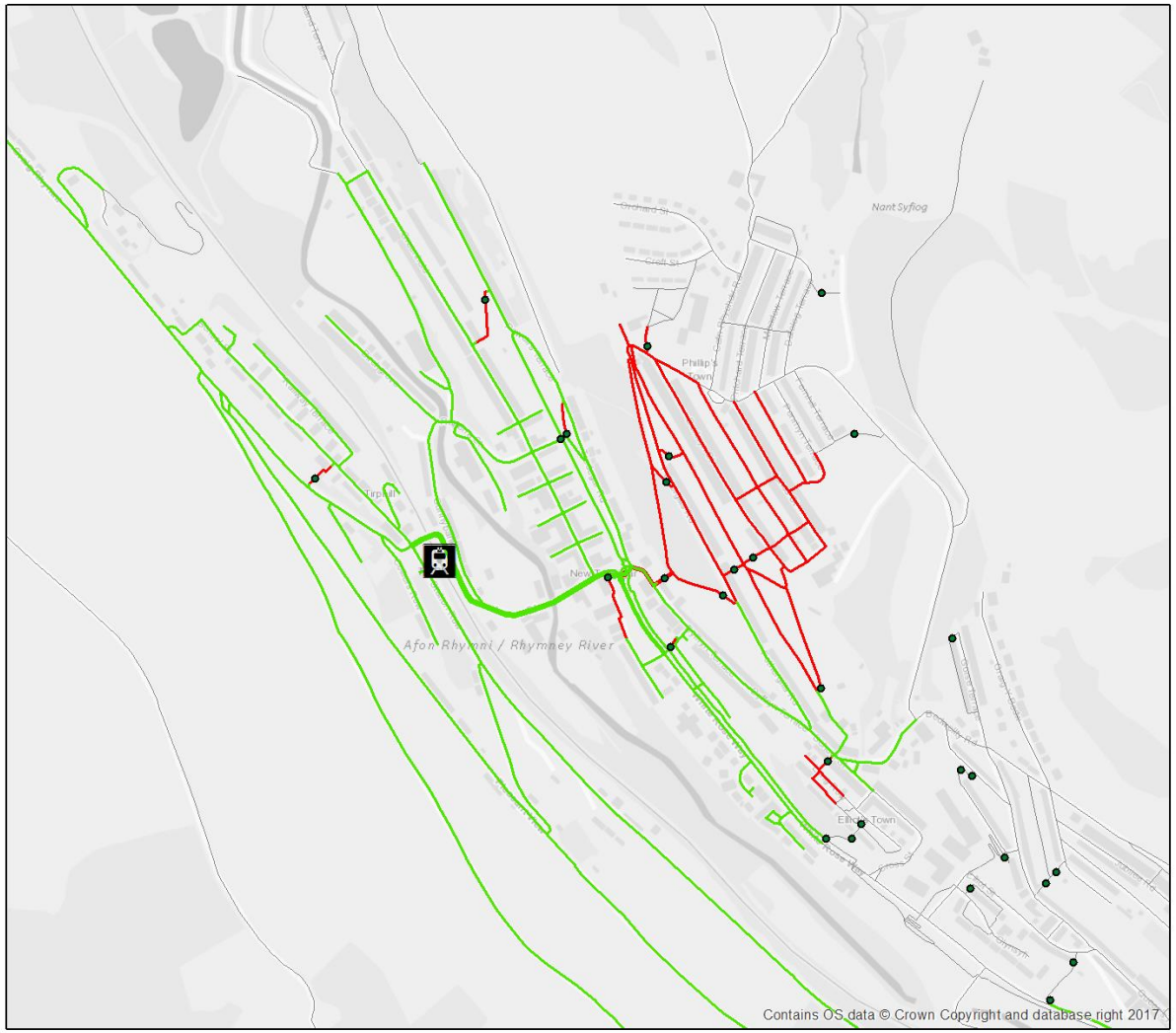
800m accessible via steps

Pilot project – not for planning use

Example: New Tredegar

Stepped route provides essential shortcut to east, which cannot otherwise be reached within 800m

Step free route takes detour to south



0 200 Meters

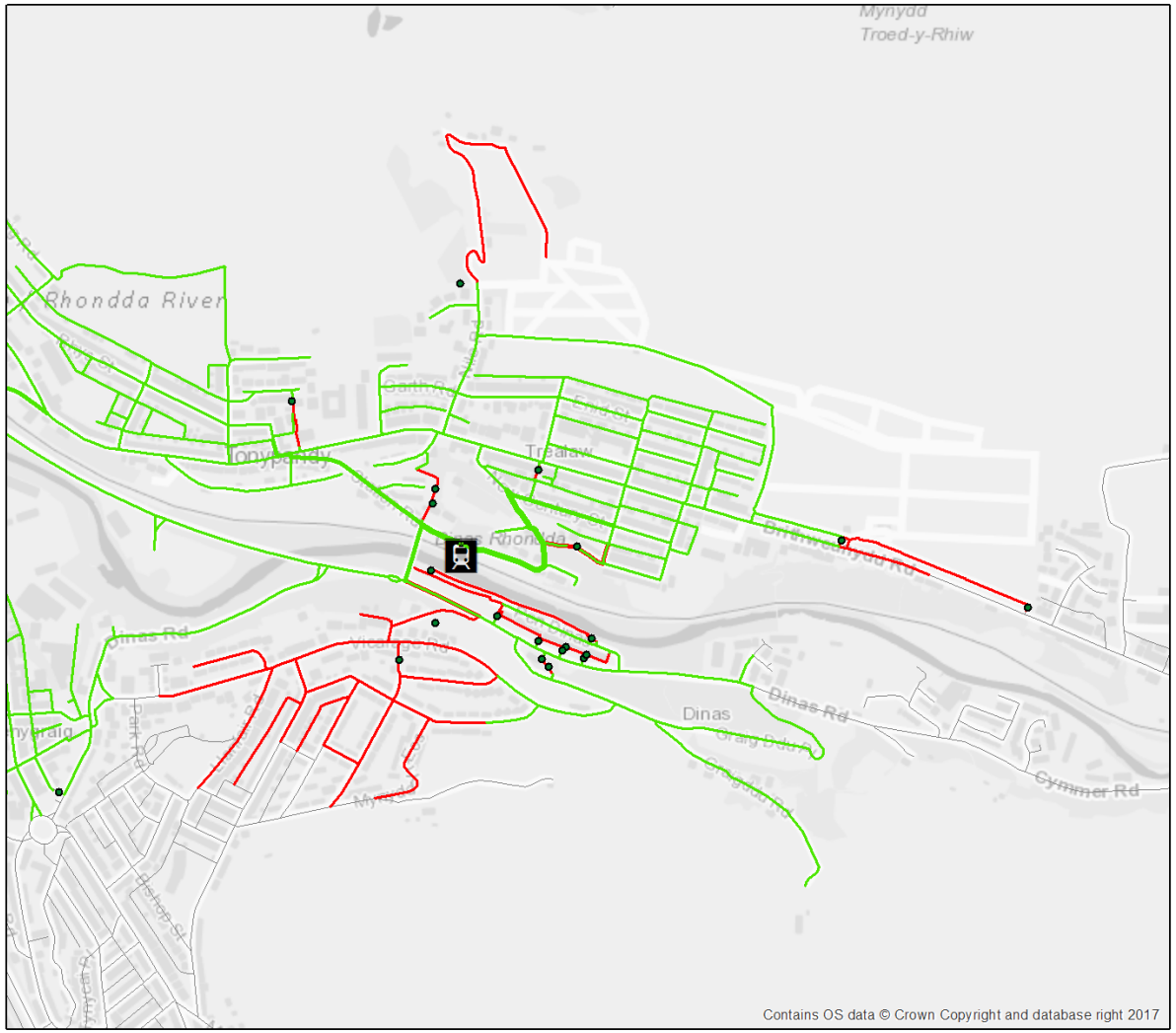
Legend

- Steps
- Rail Stations
- **800m accessible to all**
- **800m accessible via steps**

Pilot project – not for planning use

Example: Tonypandy

Stepped route provides essential shortcut to south, which cannot otherwise be reached within 800m



0 200 Meters

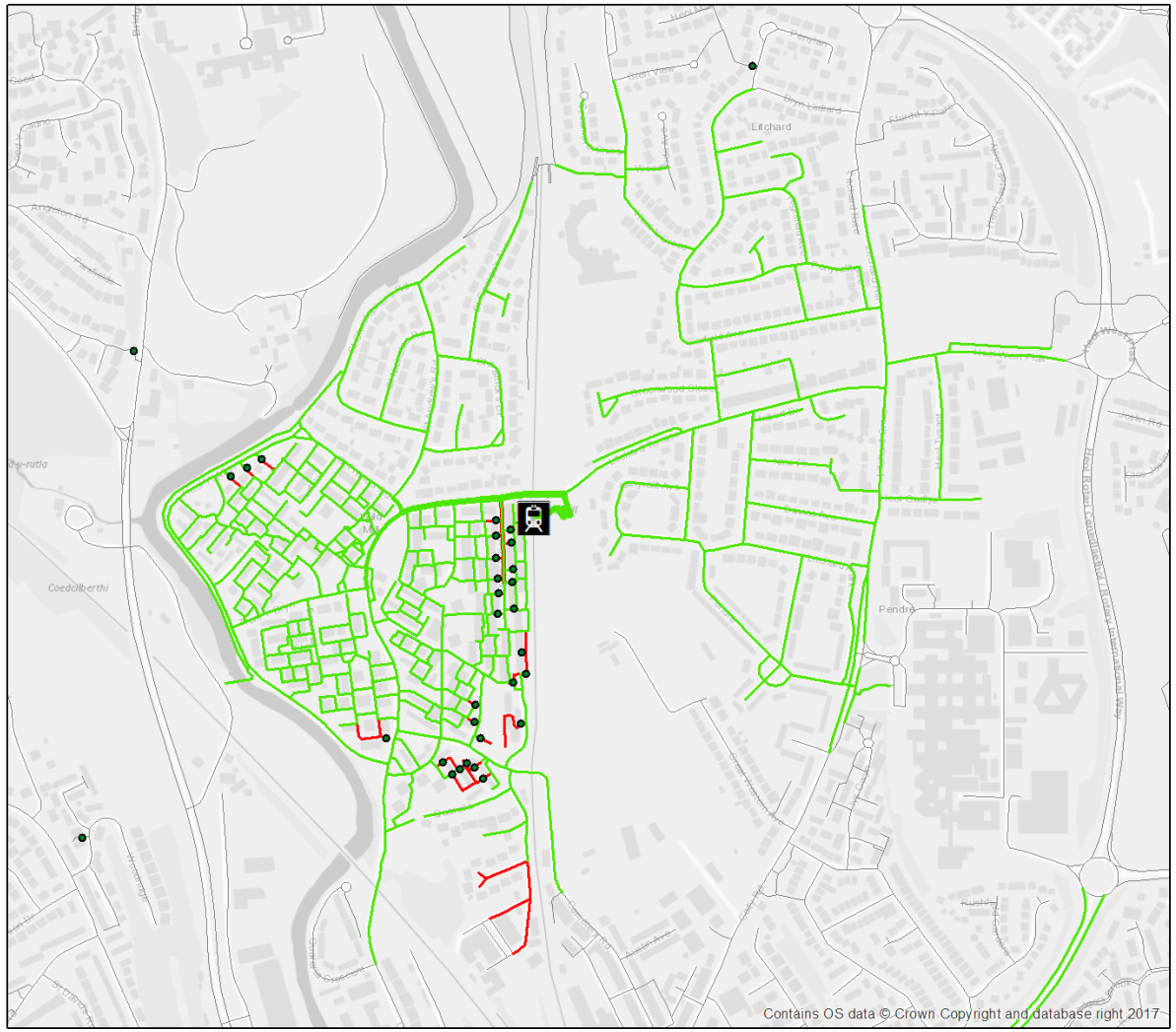
Legend

- Steps
- Rail Stations
- **800m accessible to all**
- **800m accessible via steps**

Pilot project – not for planning use

Example:
Wildmill (Bridgend)

High incidence of steps near station, but none are on essential route so impact is small



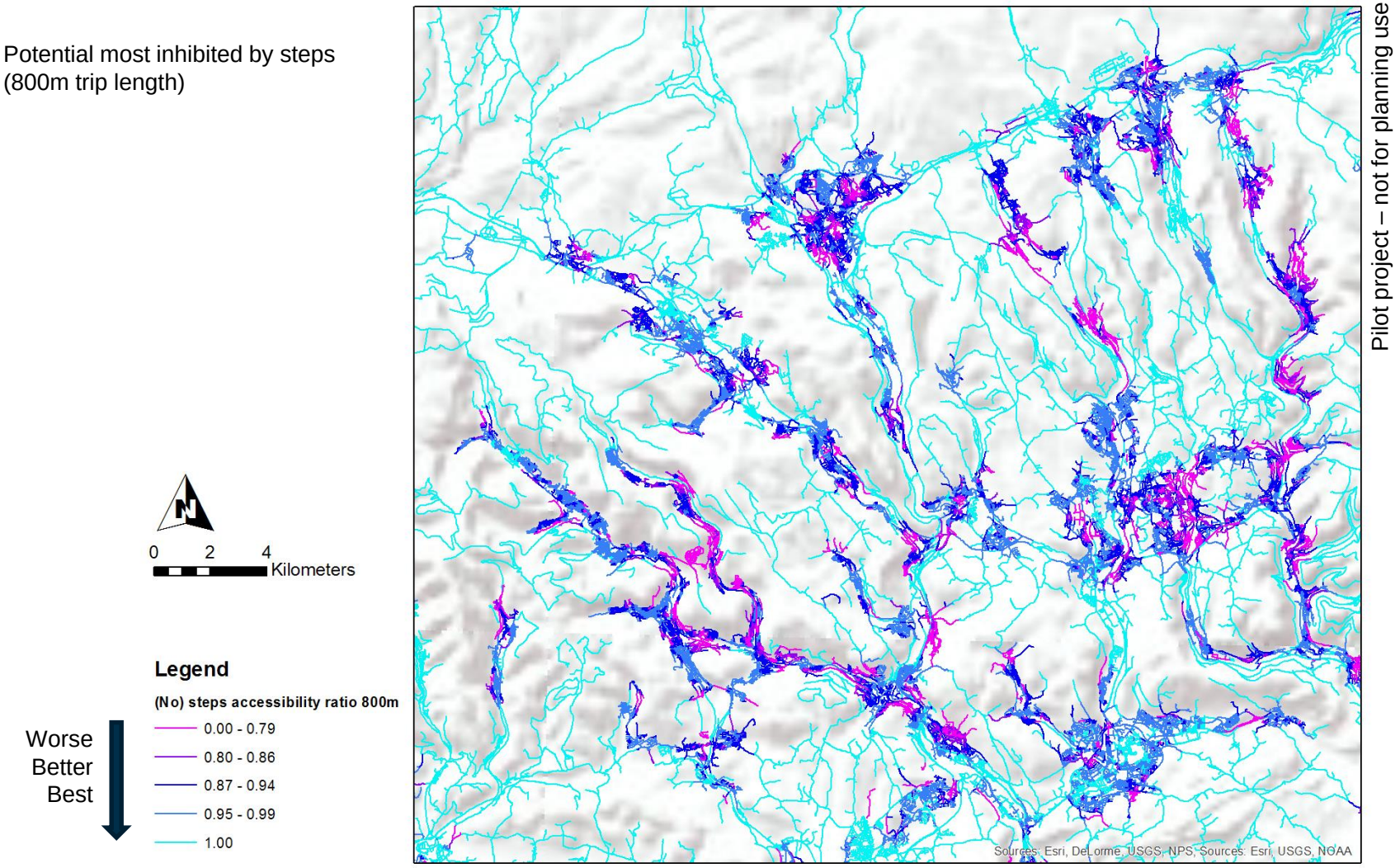
0 200 Meters

Legend

- Steps
- Rail Stations
- **800m accessible to all**
- **800m accessible via steps**

Pilot project – not for planning use

Potential most inhibited by steps
(800m trip length)



Pilot project – not for planning use

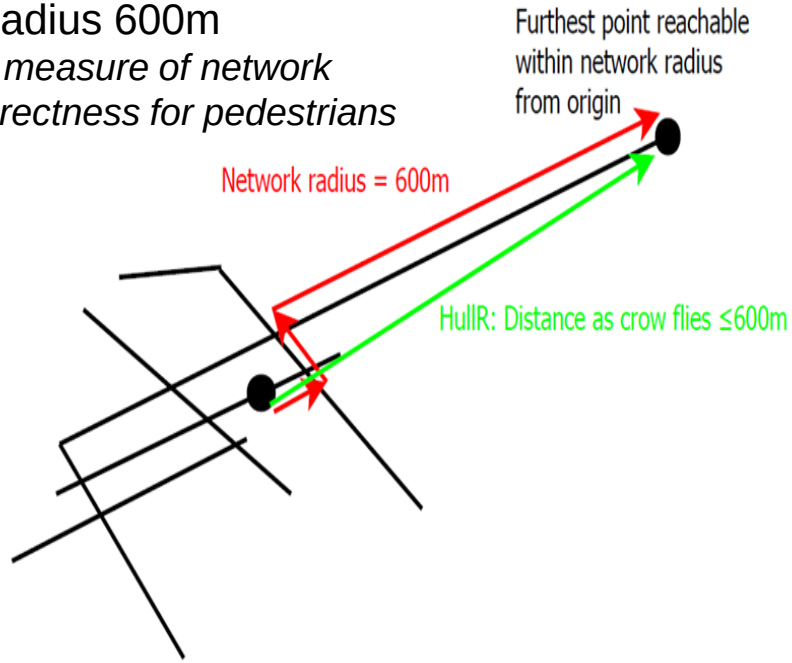
Spatially explicit measures: Convex Hull

- Spatial locality has shape, and that shape has a convex hull
  - Area
  - Perimeter
  - Shape index
  - Bearing
  - Maximum radius

Convex Hull link to Community Cohesion

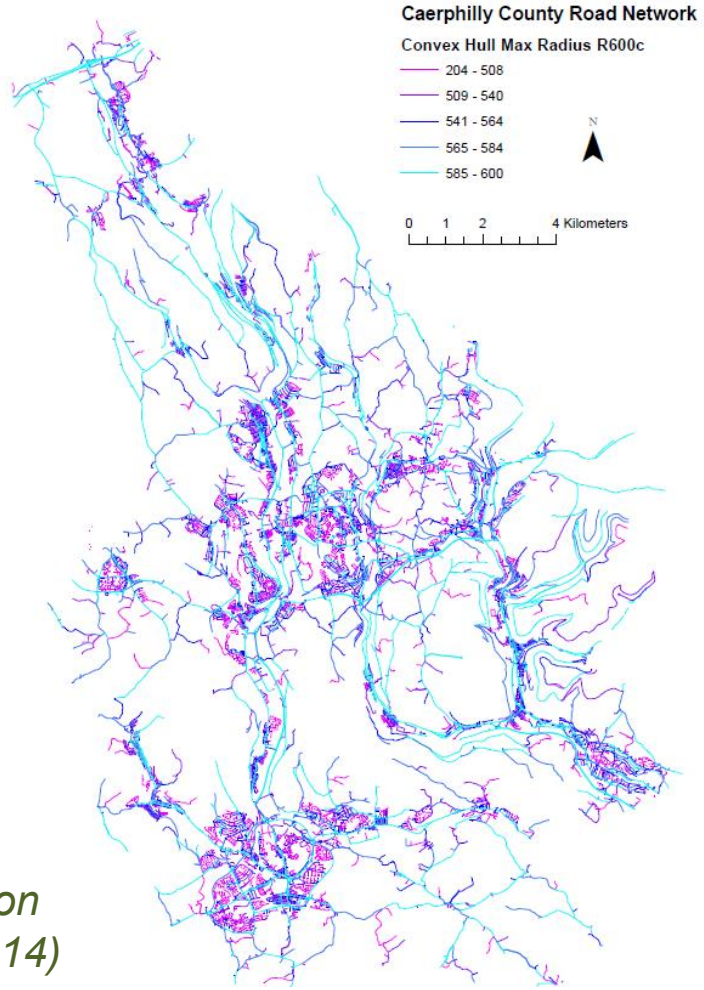
Convex Hull Maximum Radius 600m

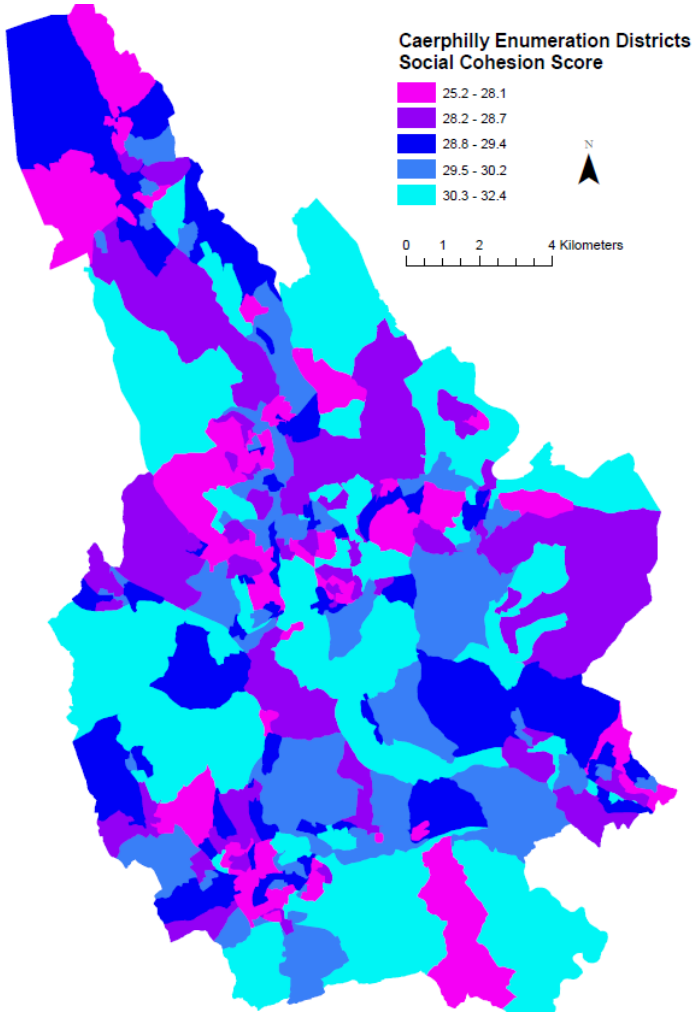
A measure of network directness for pedestrians



"the furthest you can get (as the crow flies) from your point of origin, by walking 600m (taking obstacles into account)"

Explained 10% of variance in community cohesion in 2 different studies (Cooper Fone Chiaradia 2014)





Network links to health

C. Sarkar, J. Gallacher, C. Webster, Morphometric analysis of the built environment in UK Biobank: Data analyses and specication manual (Jan 2014).

C. Sarkar, J. Gallacher, C. Webster, Urban built environment configuration and psychological distress in older men: Results from the Caerphilly

study, BMC Public Health 13 (1) (2013)

C. Sarkar, C. Webster, M. Pryor, D. Tang, S. Melbourne, X. Zhang, L. Jianzheng, Exploring associations between urban green, street design and walking: Results from the Greater London boroughs, Landscape and Urban Planning 143 (Supplement C) (2015) 112-125.

C. Sarkar, C. Webster, J. Gallacher, Residential greenness and prevalence of major depressive disorders: a cross-sectional, observational, associational study of 94,879 adult UK Biobank participants, The Lancet. Planetary Health 2 (4) (2018) 162-173.

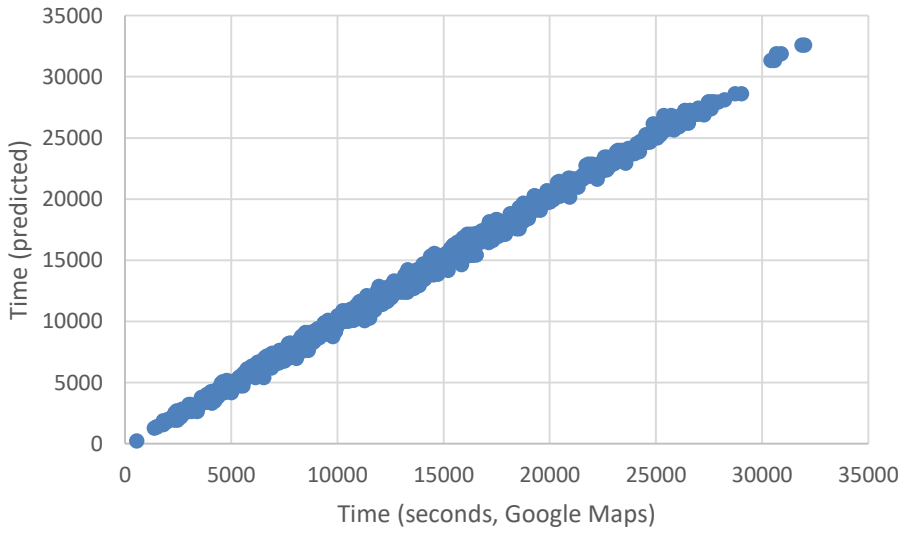
C. Sarkar, C. Webster, J. Gallacher, Neighbourhood walkability and incidence of hypertension: Findings from the study of 429,334 UK Biobank participants, International Journal of Hygiene and Environmental Health 221 (3) (2018)

- UK Biobank: large scale epidemiology project
- sDNA quantifies built environment characteristics around 500,000 homes
- Findings on depression, hypertension, mediated by walking

Network Economic links: travel time

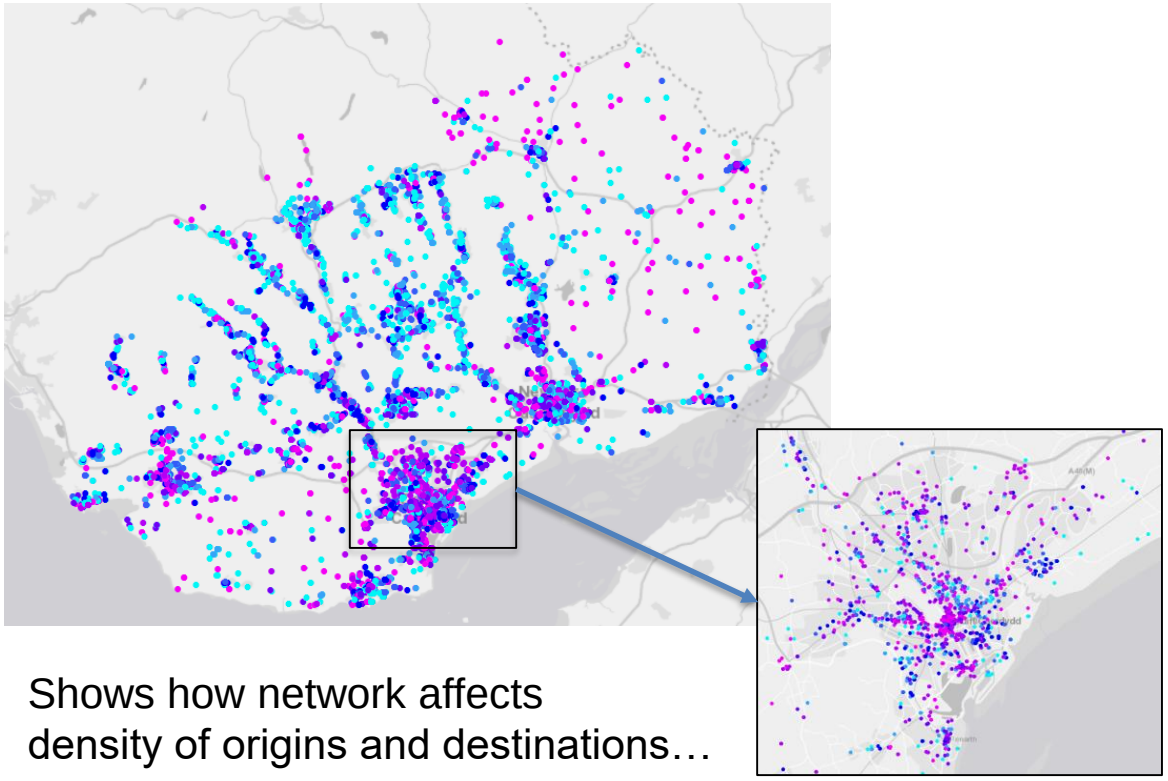
Modelling travel time from network alone

Feature	Estimated time (seconds)
1km network distance	32.0 (= 69.9 mph)
90 degree cumulative change of direction	7.61
Junction	3.75

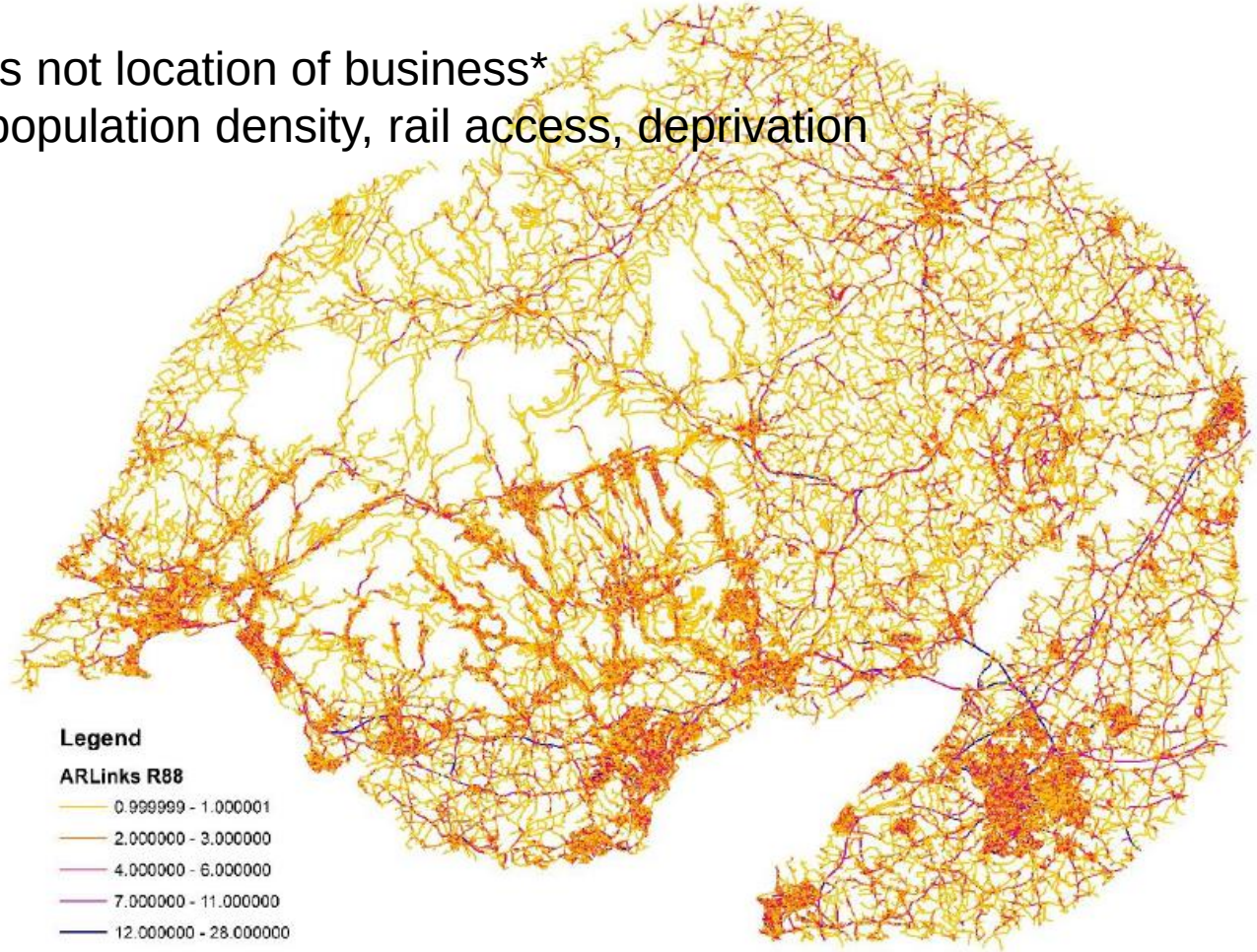


Network Economic links: Business Rates

Aamir Mohammed MSc Dissertation
Study was pitched at predicting rates for existing business not location of business*
Highly significant association (esp R880), controlling for population density, rail access, deprivation

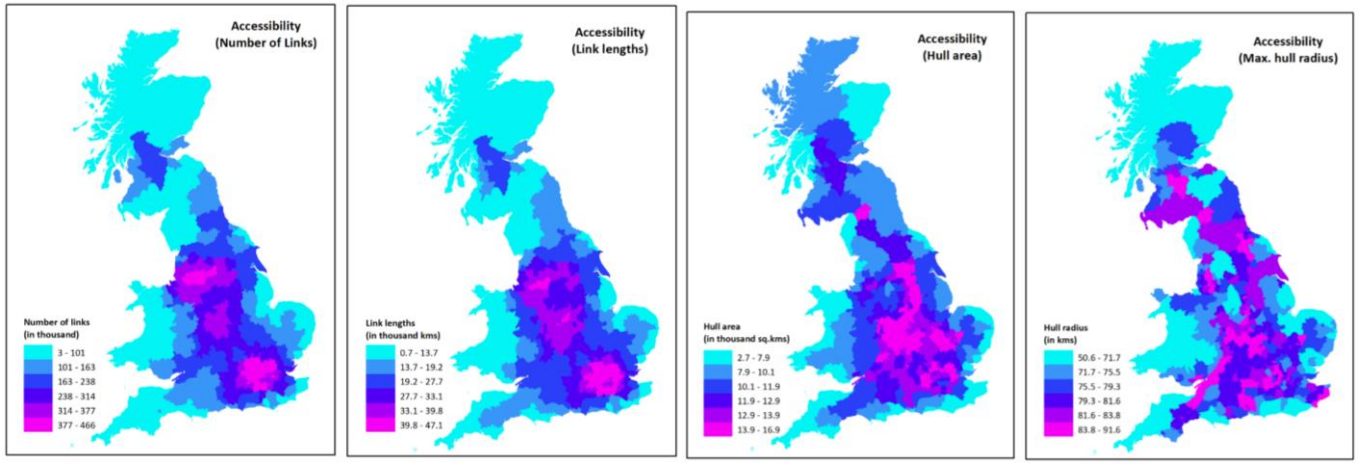


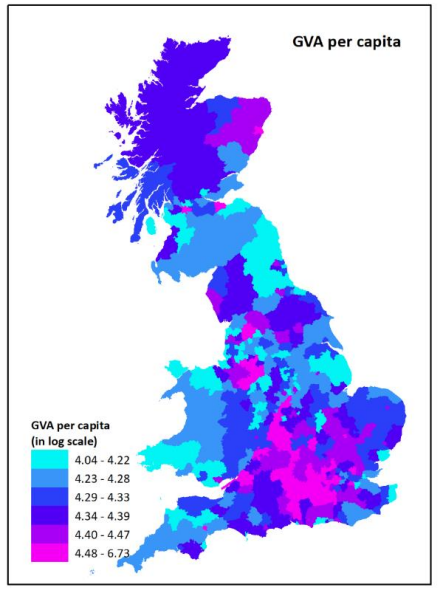
Shows how network affects density of origins and destinations…



Network Economic links: GVA

Anwar Hossein MSc Dissertation
Significant network effects on GVA (especially Convex Hull Area – an efficiency measure), controlling for population density, employment, education.

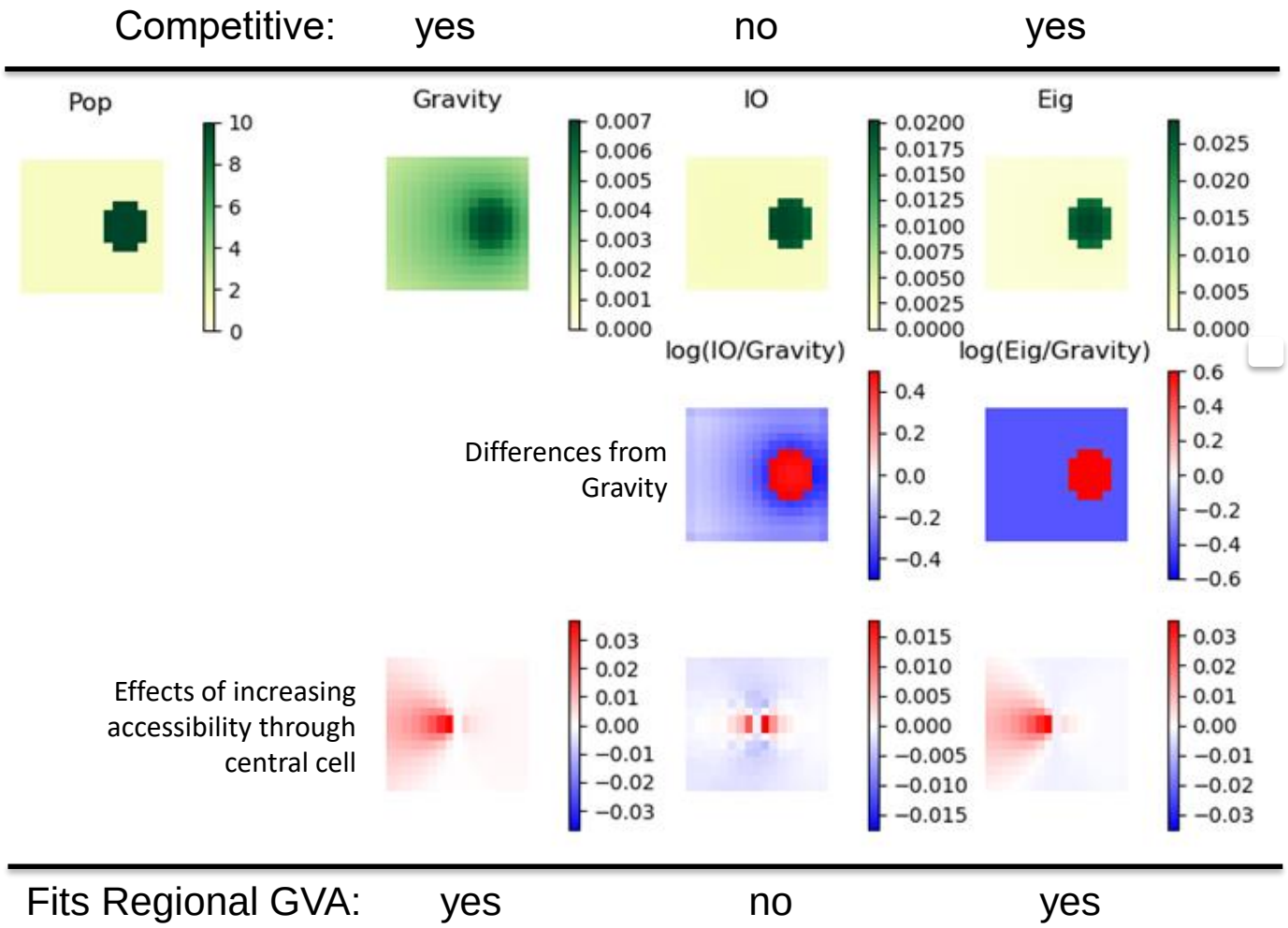




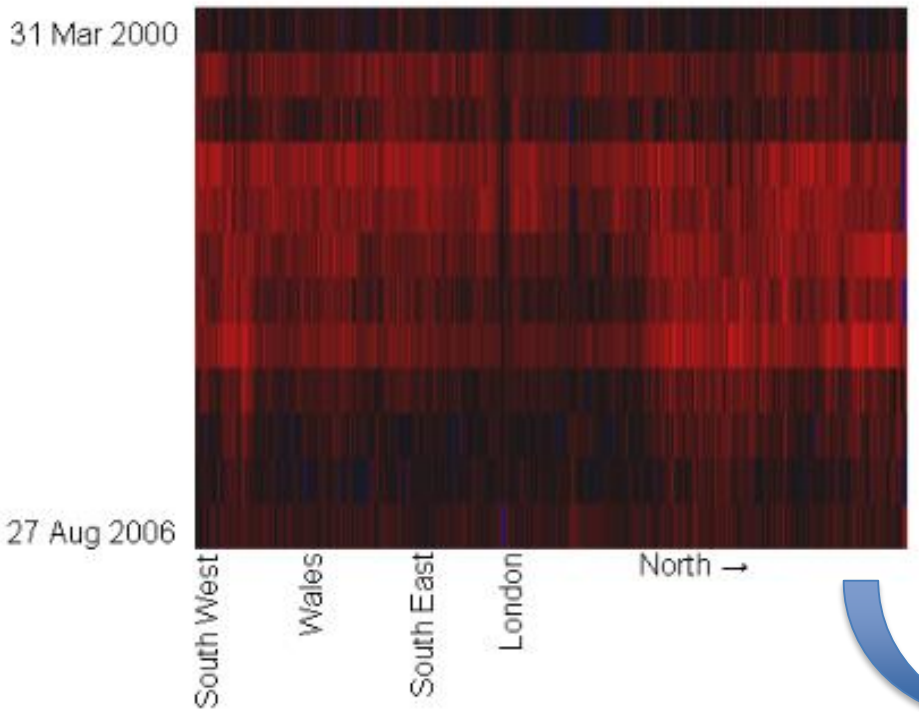
(…but, ecological fallacy. Not all business responds to accessibility in the same way; effect of accessibility on GVA appears to be both mediated and moderated by knowledge based industry (in cross sectional model. Longitudinal effects are unclear.)

Differing models of spatial competition

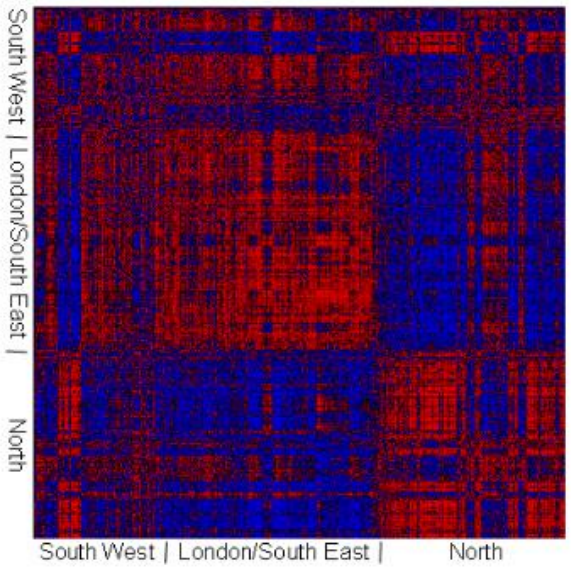
- Gravity model
  - dominant in economics, fits national GVA data
  - reducing distances always increases a gravity score, so not spatially competitive (unless used in model to divide limited resource)
- Intervening Opportunities
  - Reducing distances can reduce an IO score by diverting attention away
  - Does not fit national GVA data
- Eigenvalue centrality
  - reducing distances can reduce an eigenvalue score by diverting attention away
  - Does fit national GVA data
- Bayesian MCMC model comparison shows Gravity still better model however

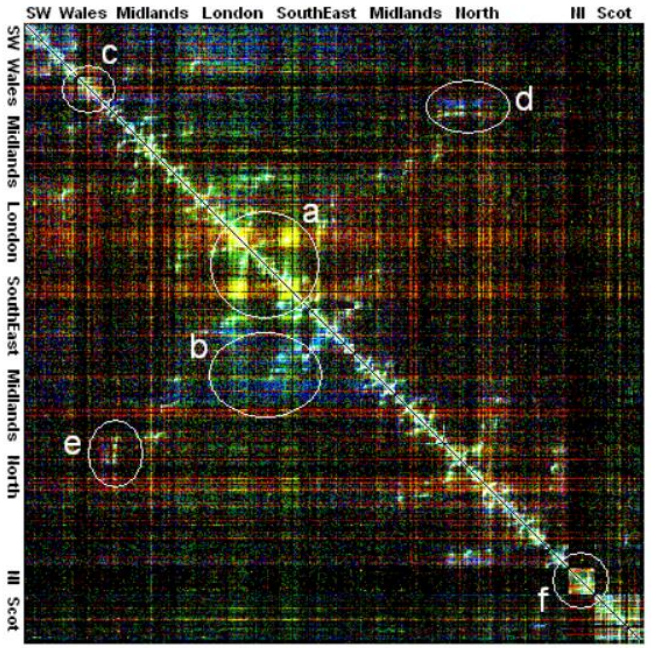


Economics: spatial time series as network



Correlation (or Granger, VAR etc)





Migration flows (ONS)

Regression

Showed ripple effect in house prices caused more by similarity between areas than spatial interaction

Cooper Orford Webster (2011)

Regionalization

- Identification of Regional Areas for the National Development Framework. Welsh Government, 2017 (Webb, Harris, Cooper, Harvey, Healey)
- 54 variables spanning themes of Wellbeing for Future Generations Act, derived from & weighted by stakeholder consultation
- Network variables included
  - Quantity of accessible natural recreational space within short drive (any)
  - Quantity of accessible natural recreational space within short drive (high quality)
  - Travel time from key population centres
  - Communities derived from interaction data (Clauset, Newman & Moore 2004)
    - Migration
    - Commuting

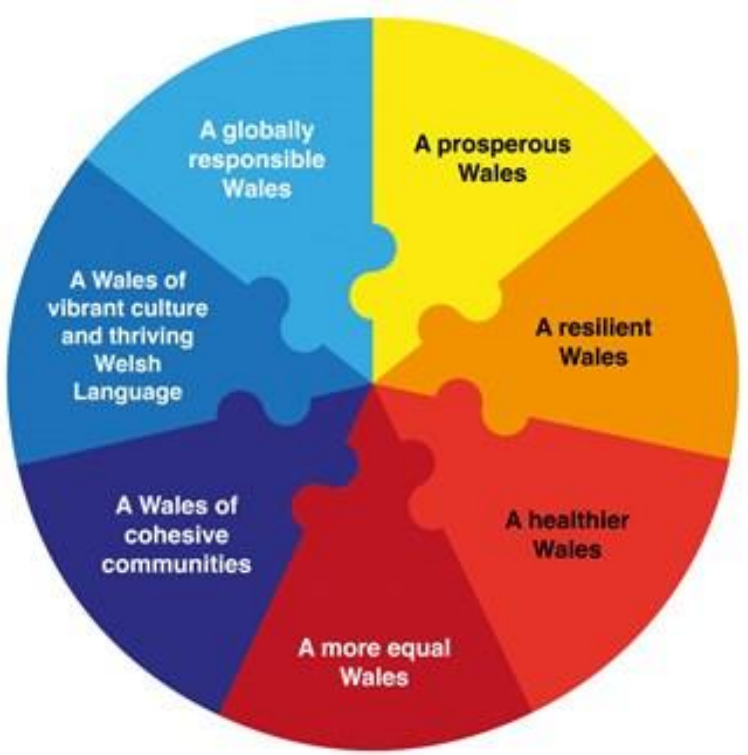


Figure 18: All Well-being Themes (Weighted) with Commuting Flows / 7 Regions

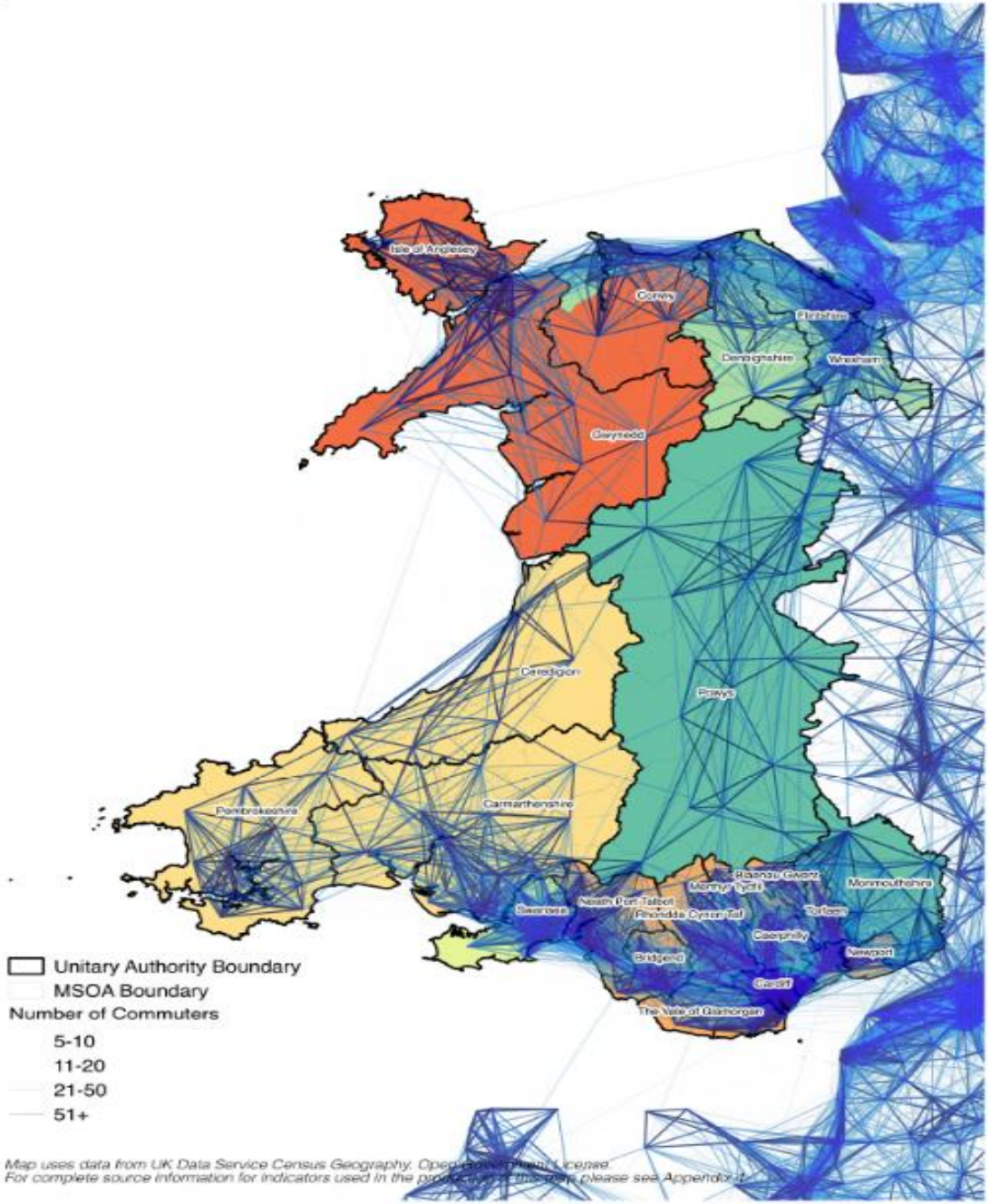


Figure 19: All Well-being Themes (Weighted) with Commuting Flows and Cross-boundary Areas / 7 Regions

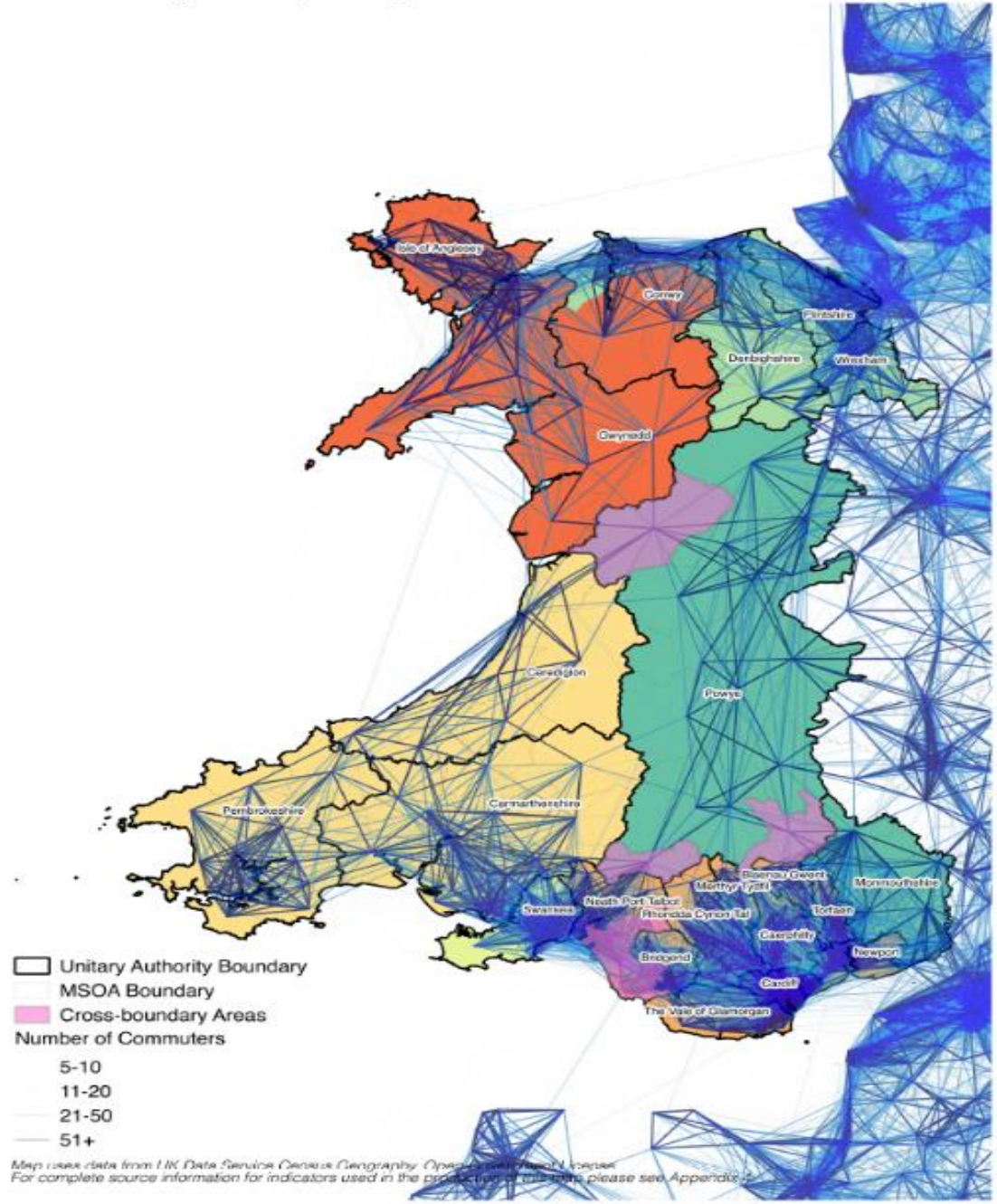


Figure 16: All Well-being Themes (Weighted) / 4 Regions

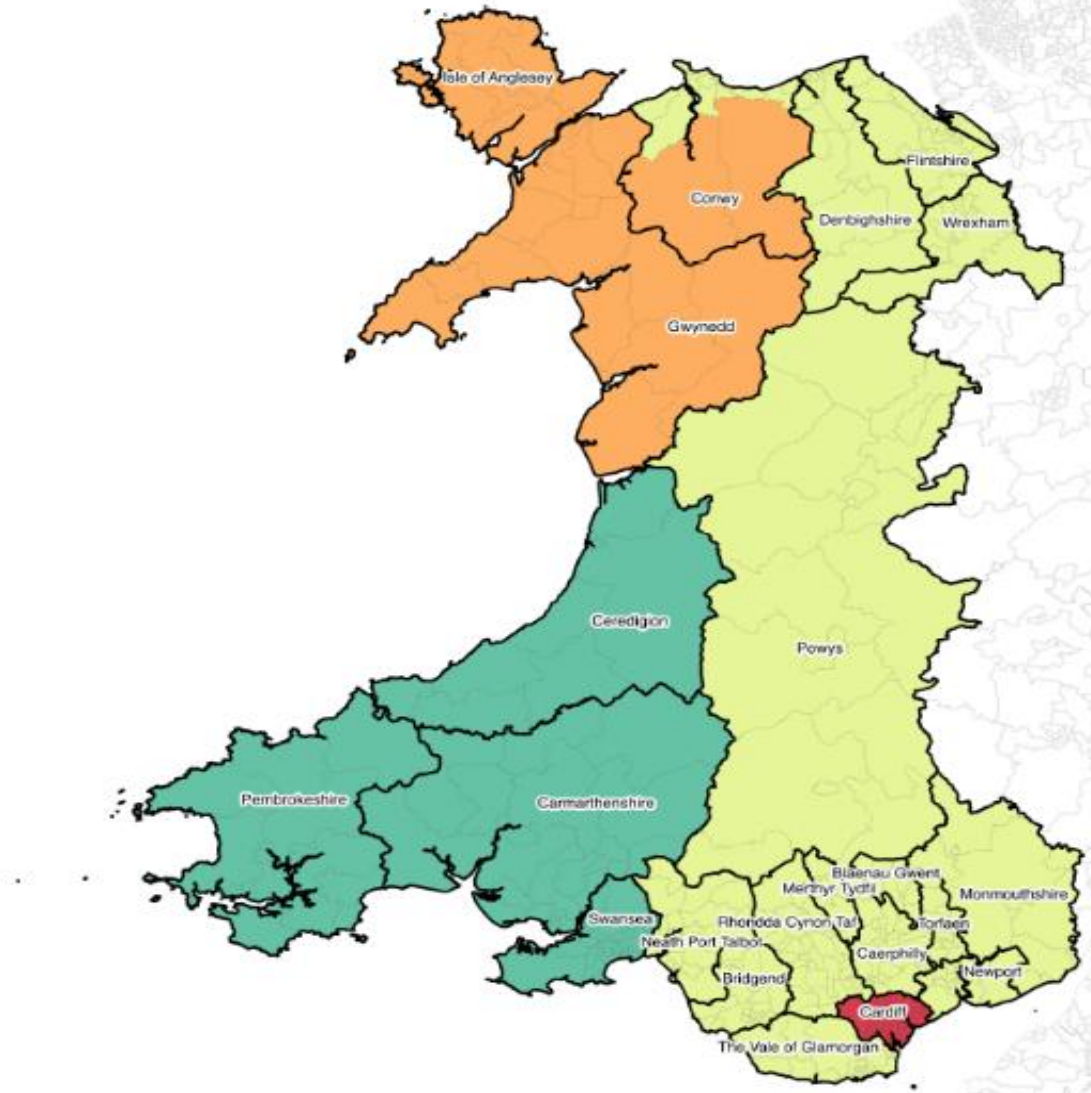
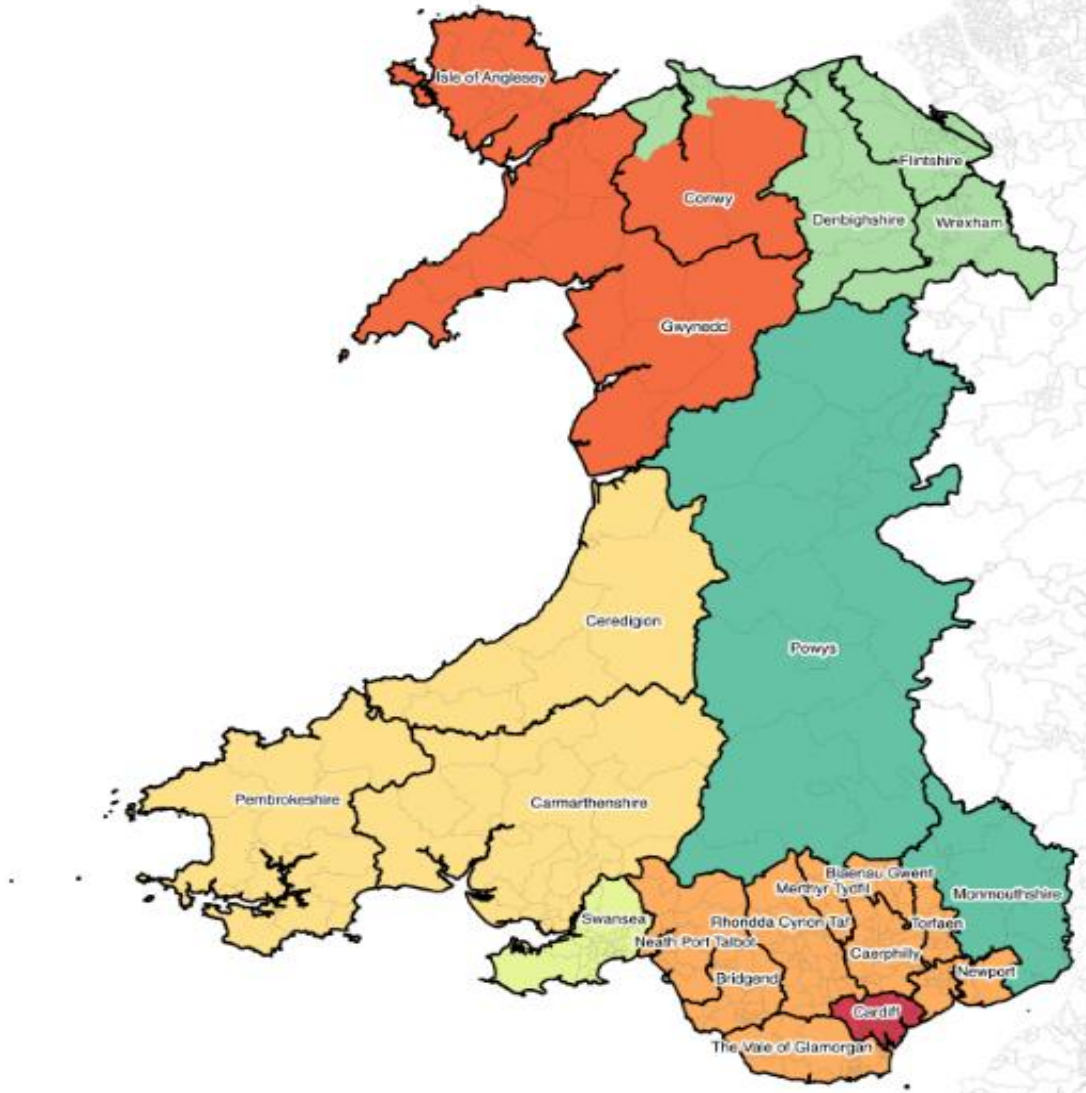
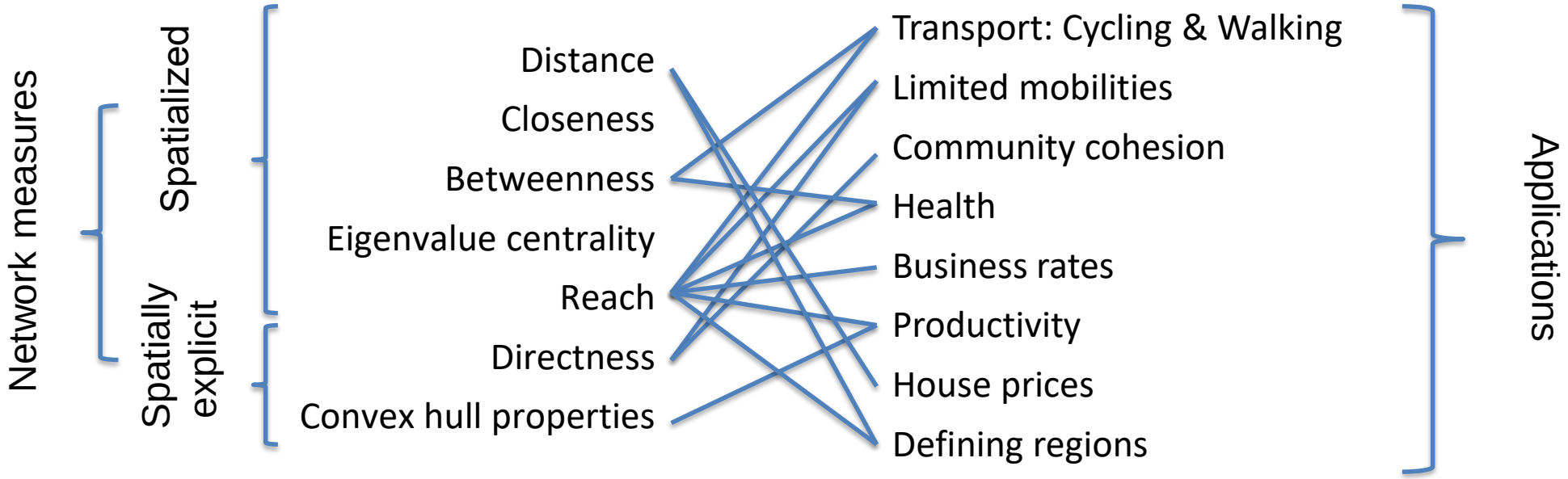


Figure 17: All Well-being Themes (Weighted) / 7 Regions



Summary



sDNA

- Don't forget the Free software!
  - Plugin for ArcGIS, QGIS, Autocad