Modelling airline competition using machine learning

Build by Vadim Skritskii
Presented By Nils G-H

INFARE

About **Infare**

Infare is the **market leader** in providing **Competitor Airfare Data** and **Pricing Intelligence Solutions** to the Travel Industry

We are trusted by **240+** Airlines

~2 Billion unique airfares collected daily

Covering 90% of commercial aircraft routes

~3.5 Billion observations distributed daily

Machine Learning & Artificial Intelligence

©2019 Infare Solutions A/S. All Rights Reserved.

INFARE

90% of Infare´s data was generated over the **last 3 years**

Our biggest **Airline Customers** have increased their **data consumption by 20 times** over the **last year**

Yet, we still only deliver **0,3% of all Airfare searches** back to airlines for competitor intelligence

Machine learning modeling of airline competition

INFARE

Motivation

Offer Airlines a better use of their searches, as they are not unlimited, nor cheap taking Total Cost of Accusation into account

Historically, airlines competition were seen from an origin and destination POW, or city codes view (and in many cases still are, as far as we observe)

Nowadays, city code view is not expressive enough to grasp complexity of competitive situation

It is widely agreed that c**ompetition is a driving force in the airline industry,** why benchmarking against the relevant competition at anytime is essential

To match how travellers see airline competition, we initiated to research into airline competition from a geographic perspective.

Geography includes: **GPS coordinates, distances associated with them, cities and their population, distances and travel time between cities and airports.**

Over the geographic level **airline capacity has been added.**

Competition Algorithm’s Rationale

To mine market intelligence, we decided to analyze airline competition from a geographic perspective

GPS coordinates, cities and their population, distances & travel time between cities and airports

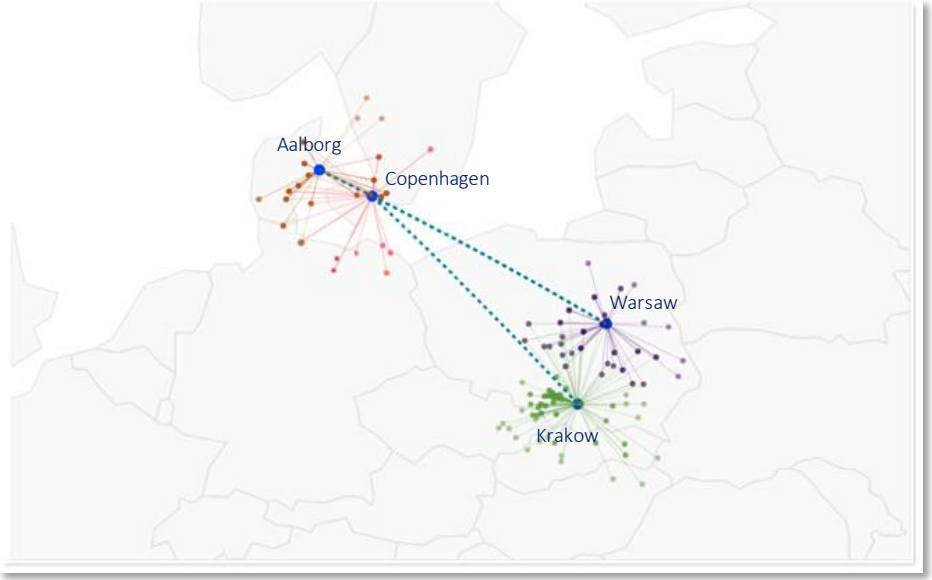
Airline capacity data added on top of the geographic layer

©2019 Infare Solutions A/S. All Rights Reserved.

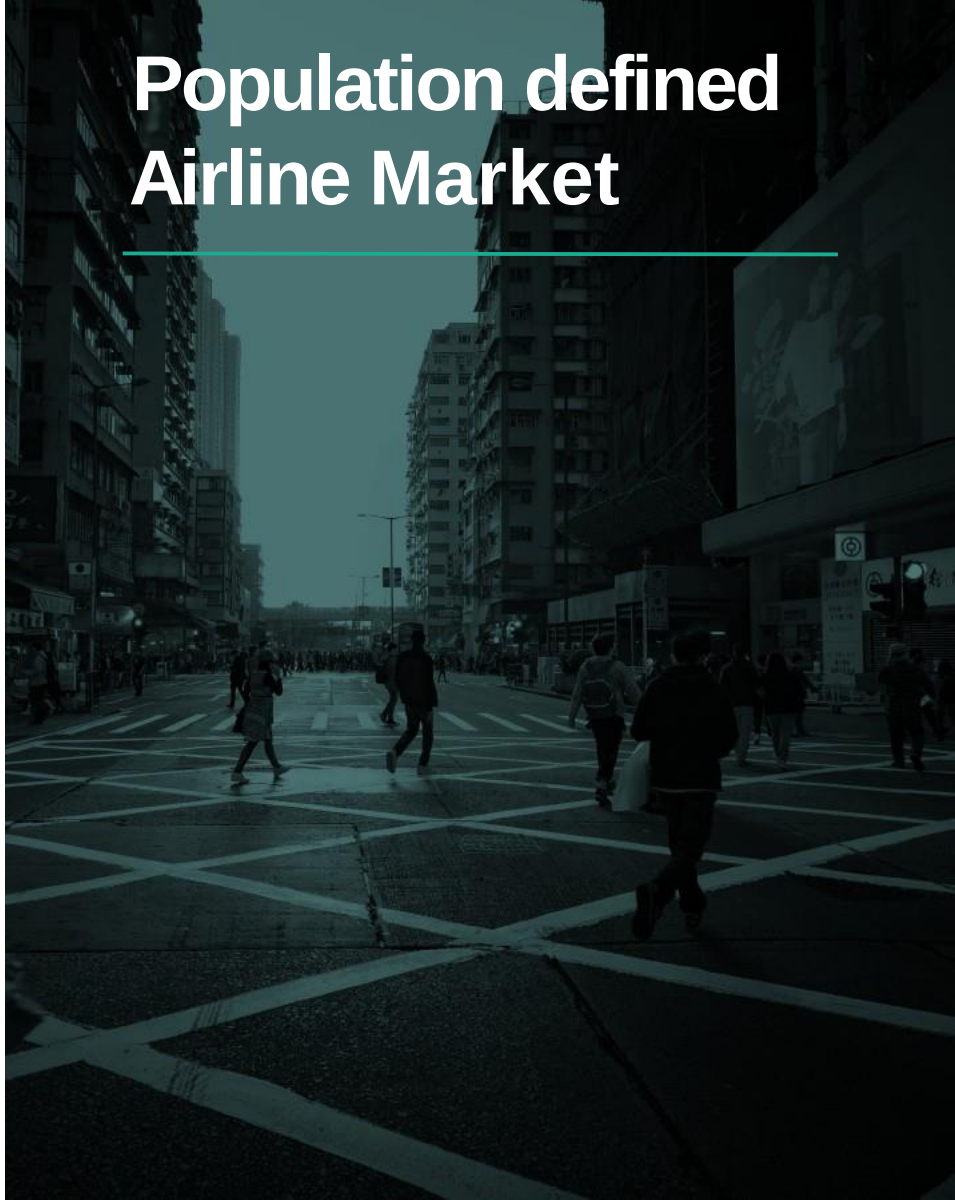
INFARE

INFARE

Population defined Airline Market

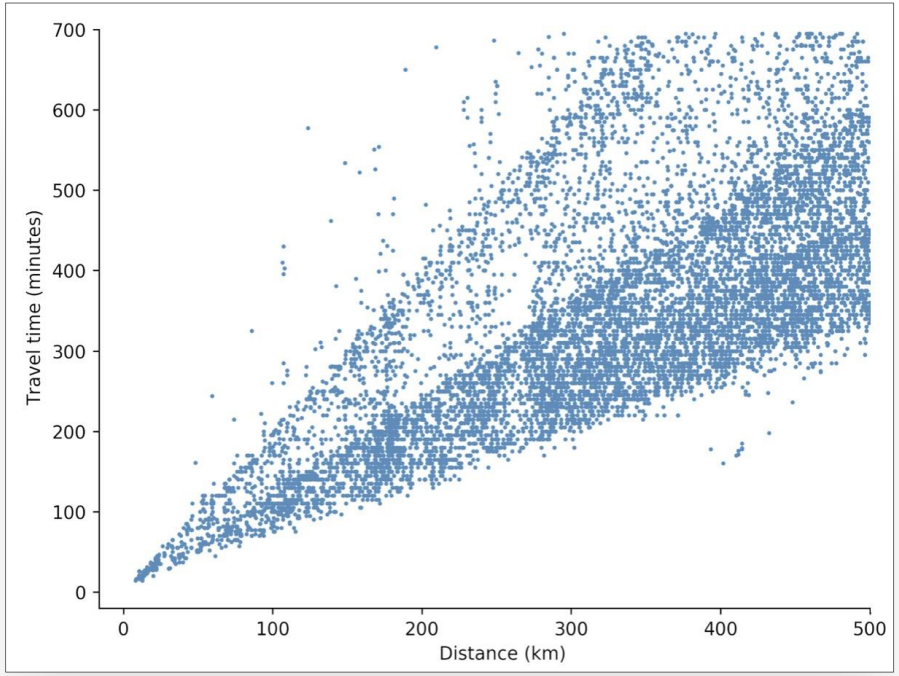


Population-defined Airline Market

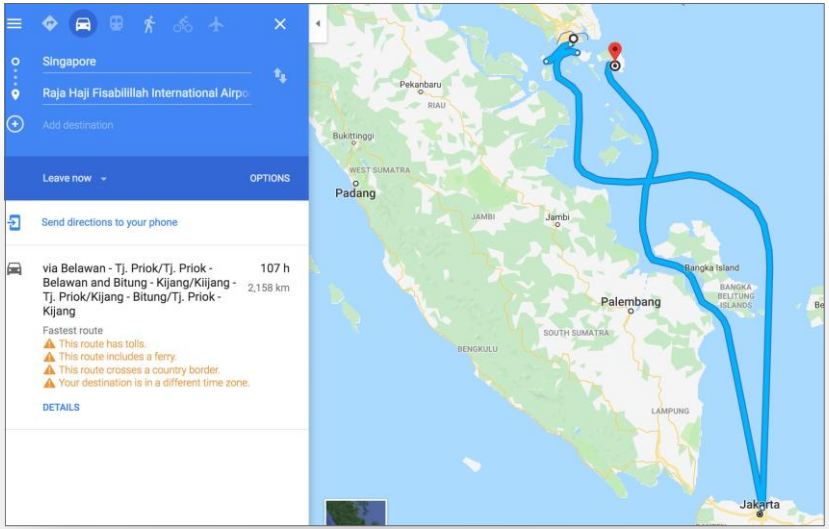


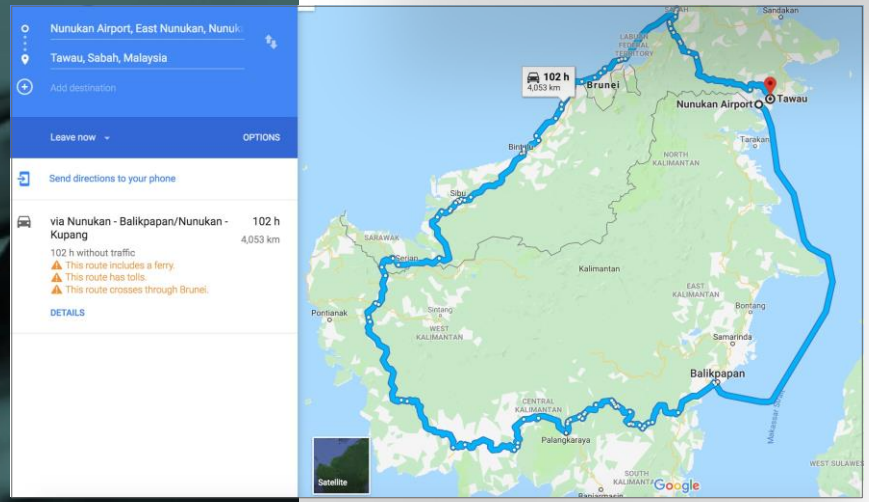
INFARE

Travel time instead of distance



Travel time more reliable than straight line distance





INFARE

INFARE

Competition algorithm

Key Points

1. Use of additional data sources bring new dimensions. (Catchment marked, and network capacity (by Innovator)

3. Compare customers network, against what they are asking us to monitor today, and then against what our model suggest they should consider monitoring, to result in a suggested competitor monitor list

4. As network data including capacity is added to the model on a continued based, new suggested list is generated by the model as it runs on a daily basis, also taking changed subscription into account

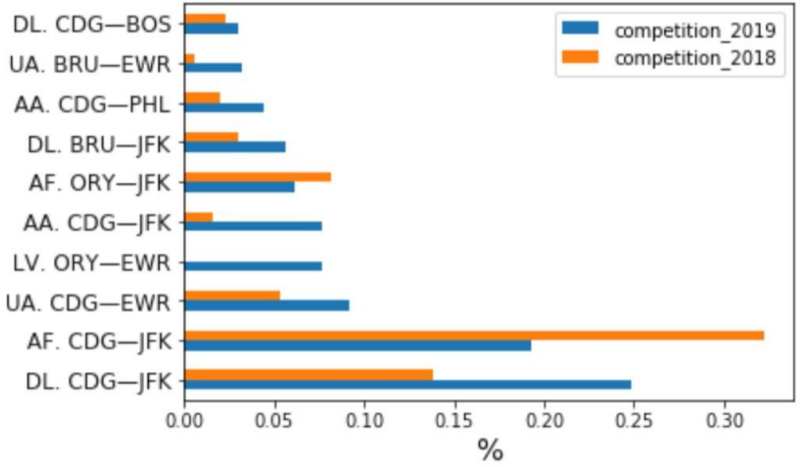
5. Marginalize two airport or city competition against the entire catchment area

6. Fine tune and validate predictive power ML based on new data flow

7. Result: Transparent and interpretable results, including suggestions

INFARE

Air France Paris to New-York market positioning: Stage 1



Dynamic Route Monitor

Used today to advise customers: Stage 2

The recommendations (*Suggestions*) are always based on the Airline's Current Subscriptions

INFARE

Dynamic Route Monitor

XYZ

Selected Profile

Profile ID

XYZ

OD

KRKCPH

KRKCPH

Select Feature

- Competition
- Seats

Filter Competition Score

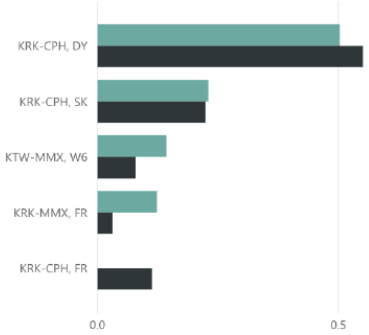
0.01 0.89

Filter Importance

- average
- high

No seats 01Jan to 31Mar

Attribute competition_2018 competition_2019



Current subscriptions

Origin	Destination	Booking site	Carrier searched	Search class	Min stay	One way	Pattern	Frequency
KRK	CPH	Norwegian	DY	E	0	1	1-180	1xDaily
KRK	CPH	Scandinavian Airlines	SK	E	0	1	1-180	1xDaily

Suggestions

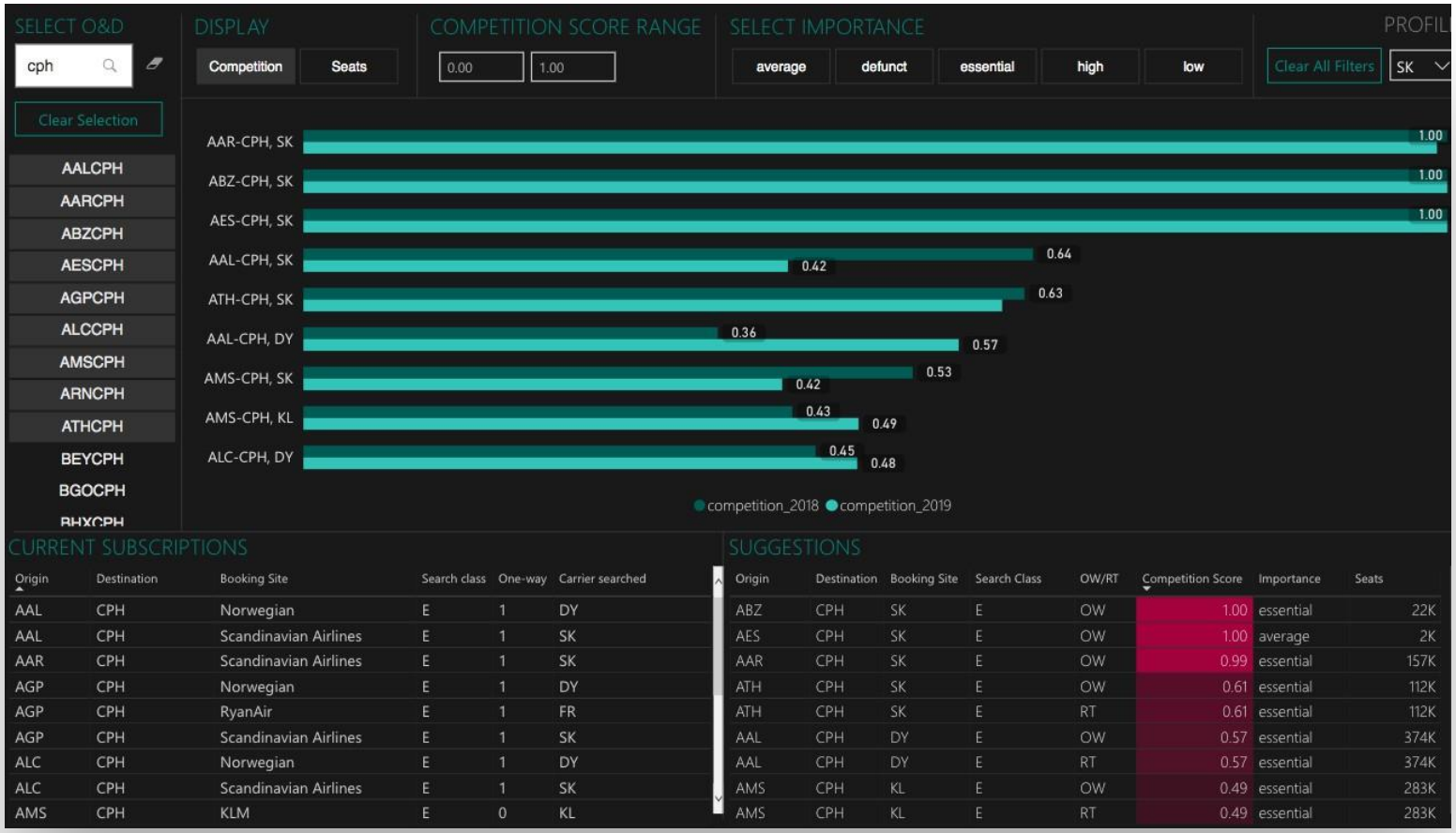
Competition Score	Seats	Importance	Origin	Destination	Booking site	Search Class	Min Stay	Pattern	OW/RT	Frequency
0.11	8694	average	KRK	CPH	FR	E	MN0	1-180	OW	Daily
0.08	15300	average	KTW	MMX	W6	E	MN0	1-180	OW	Daily
0.03	6048	average	KRK	MMX	FR	E	MN0	1-180	OW	Daily

Should an Airline not be a customer, the Suggestions will be made purely on the basis of market competition data

©2019 Infare Solutions A/S. All Rights Reserved.

INFARE

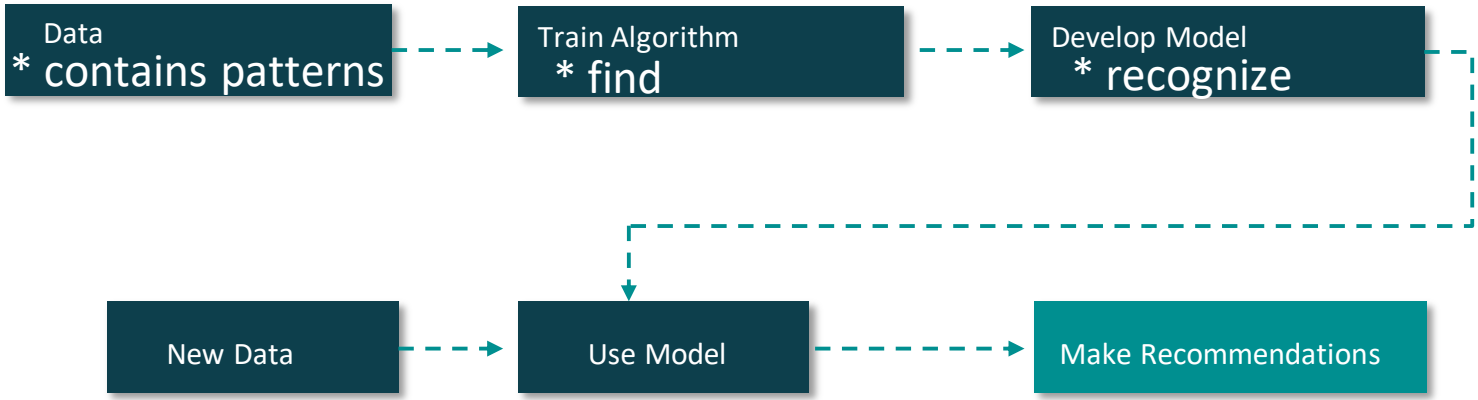
Stage 3: This version assist customers determine who the most relevant competitors are on an ongoing basis



Origin	Destination	Booking Site	Search class	One-way	Carrier searched
AAL	CPH	Norwegian	E	1	DY
AAL	CPH	Scandinavian Airlines	E	1	SK
AAR	CPH	Scandinavian Airlines	E	1	SK
AGP	CPH	Norwegian	E	1	DY
AGP	CPH	RyanAir	E	1	FR
AGP	CPH	Scandinavian Airlines	E	1	SK
ALC	CPH	Norwegian	E	1	DY
ALC	CPH	Scandinavian Airlines	E	1	SK
AMS	CPH	KLM	E	0	KL

Origin	Destination	Booking Site	Search Class	OW/RT	Competition Score	Importance	Seats
ABZ	CPH	SK	E	OW	1.00	essential	22K
AES	CPH	SK	E	OW	1.00	average	2K
AAR	CPH	SK	E	OW	0.99	essential	157K
ATH	CPH	SK	E	OW	0.61	essential	112K
ATH	CPH	SK	E	RT	0.61	essential	112K
AAL	CPH	DY	E	OW	0.57	essential	374K
AAL	CPH	DY	E	RT	0.57	essential	374K
AMS	CPH	KL	E	OW	0.49	essential	283K
AMS	CPH	KL	E	RT	0.49	essential	283K

A shift towards
a more dynamic Model



INFARE

Dynamic Route Monitoring
Summery: Questions or comments?

Practical application with the objective to bring more dynamic view on competition

Competition score is interactive, **brings transparency in a complex industry.**

On top of competition algorithm a recommendation engine has been implemented.

It alerts of changes, assigns importance to each data source, and **suggest actions.**

V1: Stage 3 is live today, as internal tool to advise our customers

Next step is **including, multi-stop as well as pricing and associated product offer information**, to determine if additional travel time on ground is worth the cheaper offer, from an airport further from ones natural point of departure and/or arrival.

= increased catchment area, defined by more competitive offer (= **new competitor**)

INFARE

Competition algorithm. Models used at present version

Mixed probabilistic flow network, where leaf (initial & terminal) edges are cities, transit edges represent airports, and vertices are directed

$$G = (V, E_{city}, E_{airport})$$

Capacity function is gravity inspired, where mass of particle$_1$ is the city population, and particle$_2$ is number of offered seats

$$C(E_{city}, E_{airport}) = \frac{1}{\text{travel time}(city, airport)^2}$$

Competition algorithm

Flow from cities is defined as capacity of given **(city, airport)** vertex divided by sum of all other vertices from this city

$$w_{cap}(cit\ y_x, airport_y) = \frac{c(cit\ y_x, airport_y)}{\sum_j c(cit\ y_x, airport_j)}$$

Intuition behind is that from an arbitrary city travellers consider all available airports in their neighbourhood, and **capacity weight is a measure of convenience.**

Competition algorithm

Multiple cities merge into airport, that is expressed as set of all previously defined vertices

$$W(airport_x) = \{w(city_i, airport_x) \mid distance(city_i, airport_x) < \Theta\}$$

where

$$\sum_i w(city_i, airport_x) = 1$$

Graph source **s** and sink **t** are cities, and **(G, c, s, t)** forms a flow network.

Competition algorithm

Flow graph is weighted by city populations, which requires a set of relative populations

$$popul_{rel}(airport_x) = \left\{ \frac{popul(city_i)}{\sum_i popul(city_i)} \mid distance(city_i,\ airport_x) < \Theta \right\}$$

Subset vertices from the space of all possibilities has been identified using locality sensitive hashing, as otherwise computational complexity would be intractable.

As well a **normalization function** is needed, that maps parabolic domain into a linear codomain

$$F_{norm}(v) = \begin{cases} 2 \cdot v & \text{if} \leq 0.5 \\ 2 - 2 \cdot v & \text{otherwise} \end{cases}$$

Competition algorithm

Using pairwise operators, norm function, and weighted a competition between two airport can be expressed in the following form

$$K(a_1, a_2) = \frac{sum(F_{norm}(W(a_1) \oslash \{W(a_1) \oplus W(a_2)\}) \circ popul(a_1))}{sum(W(a_1) \circ popul(a_1))}$$

That leads to a value in range 0 to 1, **where 0 is no competition, and 1 is a worst case competition.**