Data Science in Aviation Workshop
21th May 2014. EEC

Metrics and visualisation: a (very) practical example

Samuel Cristóbal

innaxis
innovation for
a complex world

why visualisation?

- Humans have an inherent ability to find patterns, but only if the information is presented a suitable way, usually visual information.

(e.g. face recognition vs market prices evolution)

Why visualisation?

Which route performs better?

delayed flights	route A	route B
Low traffic	0 out of 10 (0%)	1 out of 40 (2.5%)
High traffic	5 out of 40 (12.5%)	3 out of 10 (33.3%)

Why visualisation?

Which route performs better?

delayed flights	route A	route B
Low traffic	0 out of 10 (0%)	1 out of 40 (2.5%)
High traffic	5 out of 40 (12.5%)	3 out of 10 (33.3%)

Either in low and high traffic levels A outperforms B, so should A perform better than B overall?

Why visualisation?

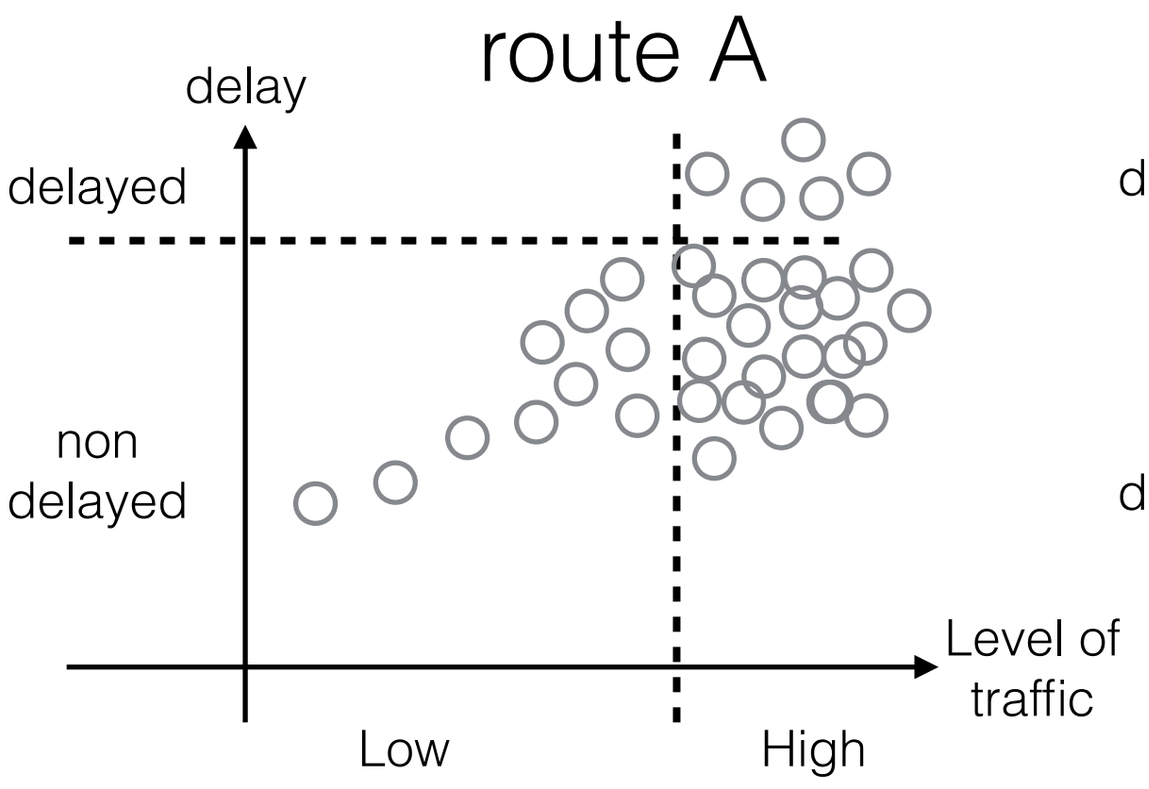
It would be temping to say that if A performs better in any scenario it should perform better overall, but…

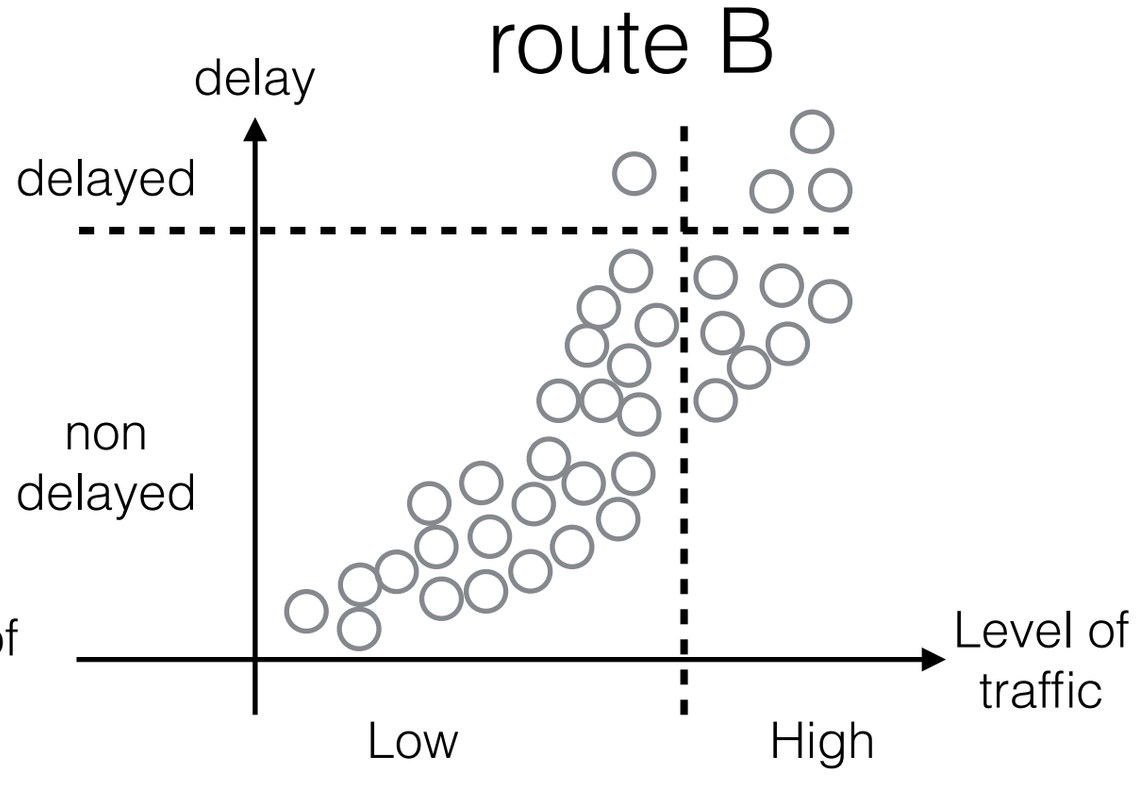
delayed flights	route A	route B
Low traffic	0 out of 10 (0%)	1 out of 40 (2.5%)
High traffic	5 out of 40 (12.5%)	3 out of 10 (33.3%)
total	5 out of 50 (10%)	4 out of 50 (8%)

So it happens that B performs better than A, overall.

Why visualisation?

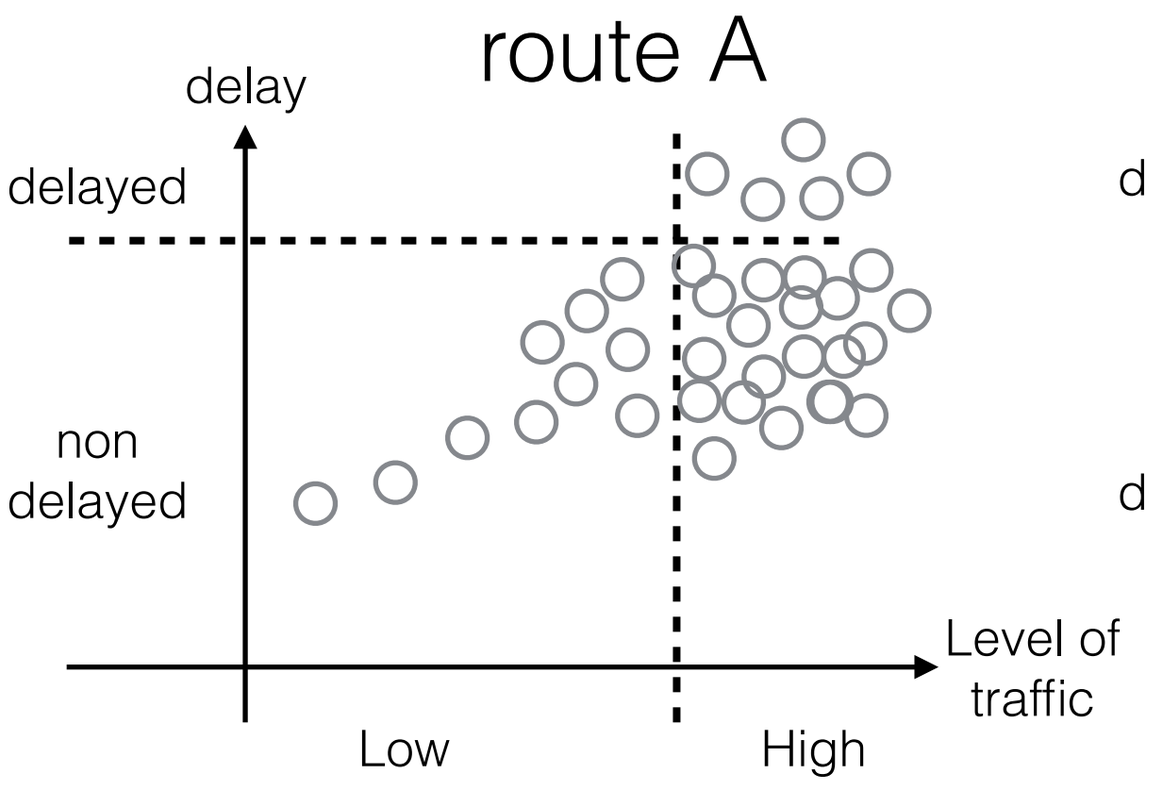
but what would have happened if I had showed you the level of traffic vs delay plots instead of the table

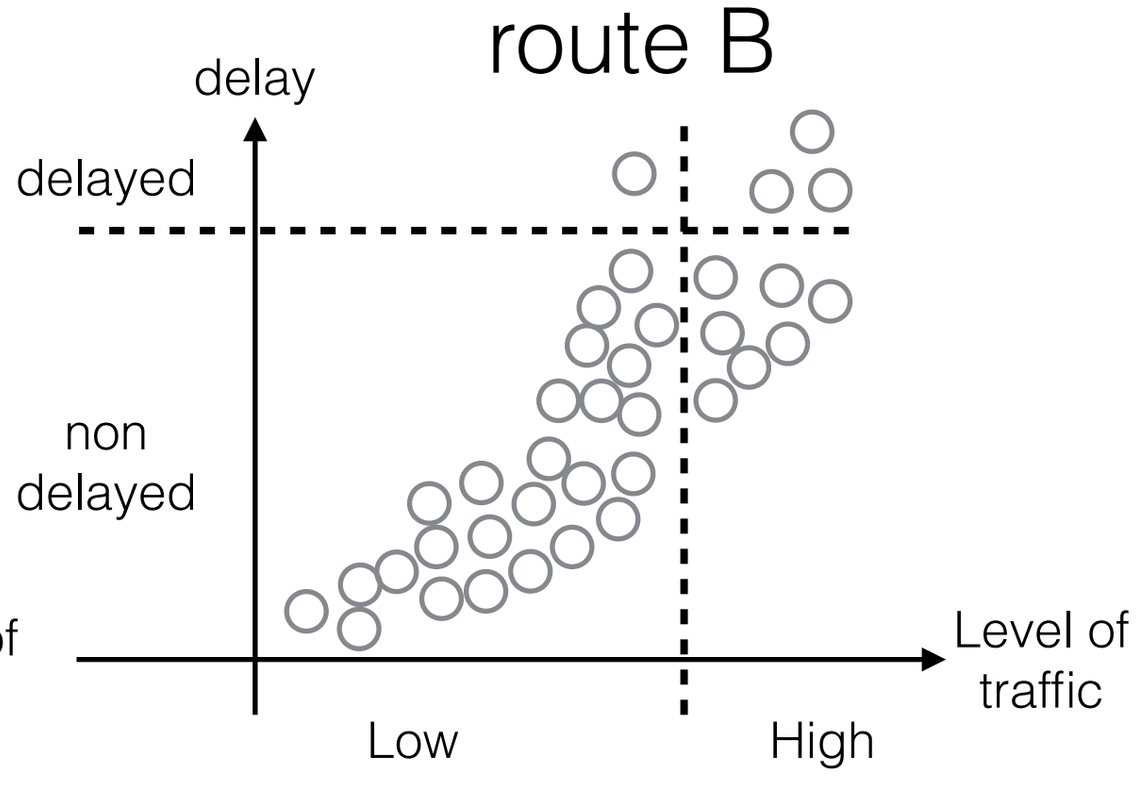




Why visualisation?

but what would have happened if I had showed you the level of traffic vs delay plots instead of the table

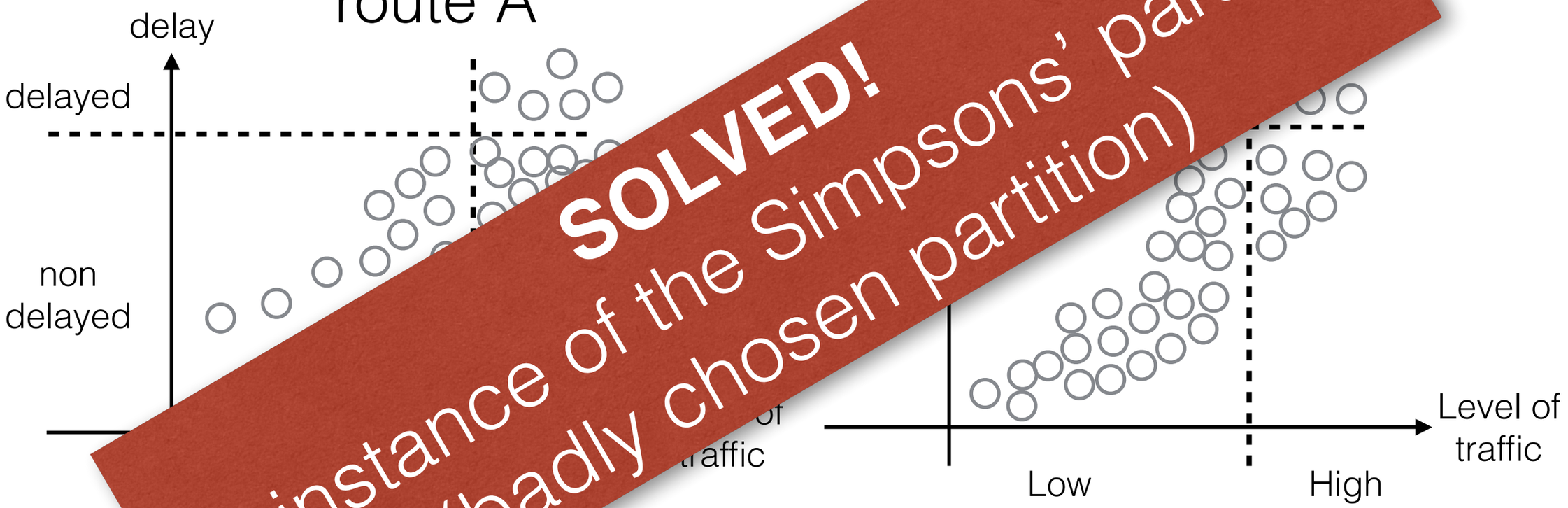




you would have instantly detected a trend in both data sets, indicating that B outperforms A

Why visualisation?

but what would have happened if I had sh ed you
the level of traffic vs delay plots inste able



SOLVED!
An instance of the Simpsons' paradox
(badly chosen partition)

you would instantly detected a trend in both data sets,
indicating that B outperforms A

the (very) practical example

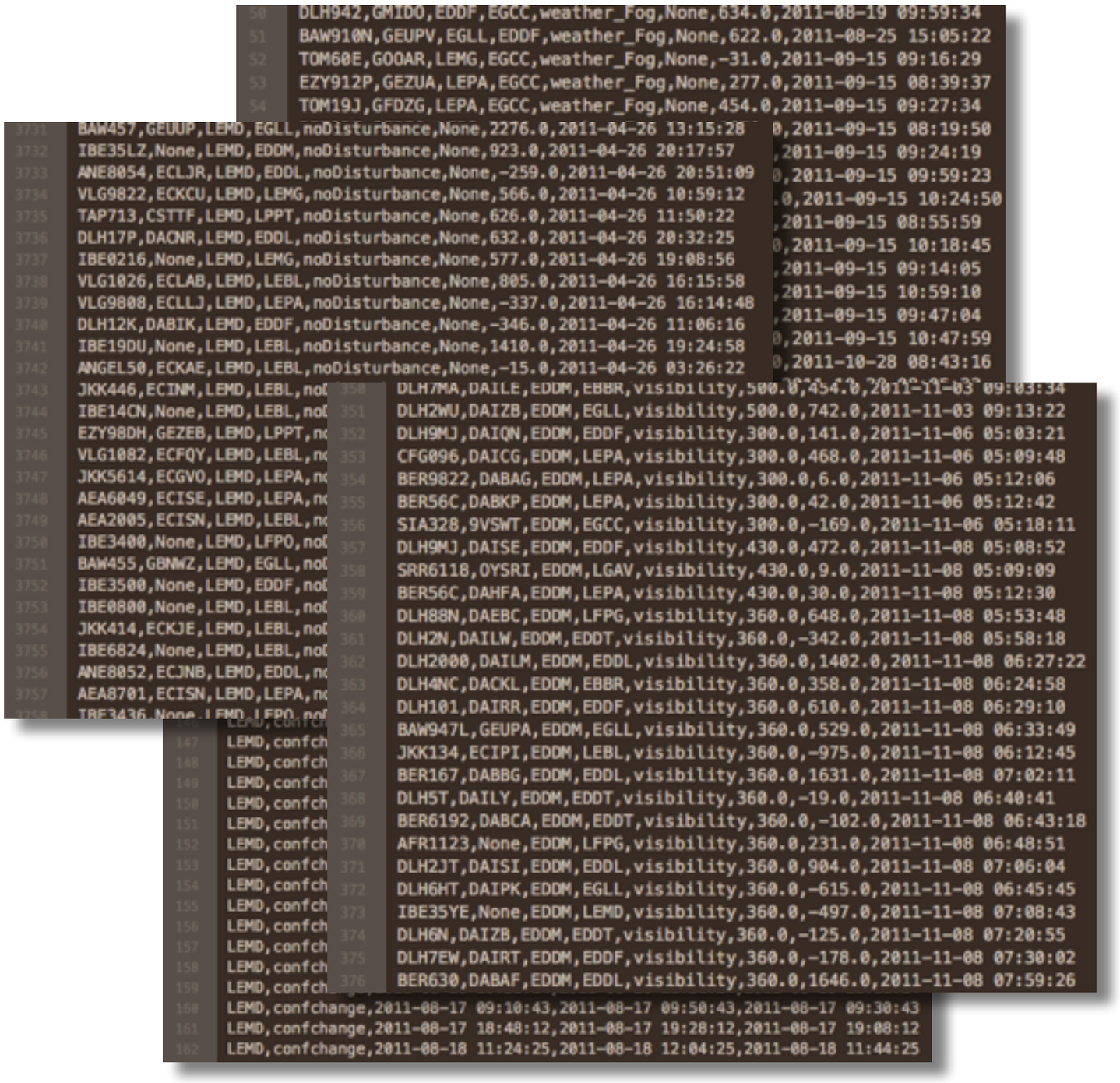
- The subsequent work has been developed in the framework of the ongoing FP7 Resilience 2050 project as an alternative metric to measure the **resilience*** of the ATM system in Europe.

*the ability to recover under abnormal conditions.

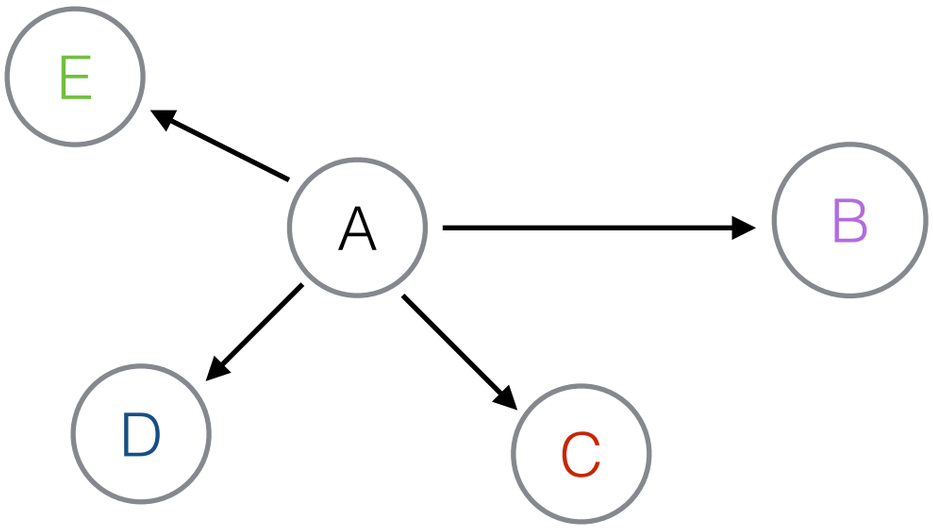
Resilience **2050**

the data set

- Nine months of traffic data within EU (March-November, 2011).
- For each flight the departure and arrival delay are known, as well as the aircraft registration and callsign.
- Additional sources of information such as Metar reports, DDR declared capacities, ACC headlines.
- Airports and geolocations have been anonymised for this presentation.

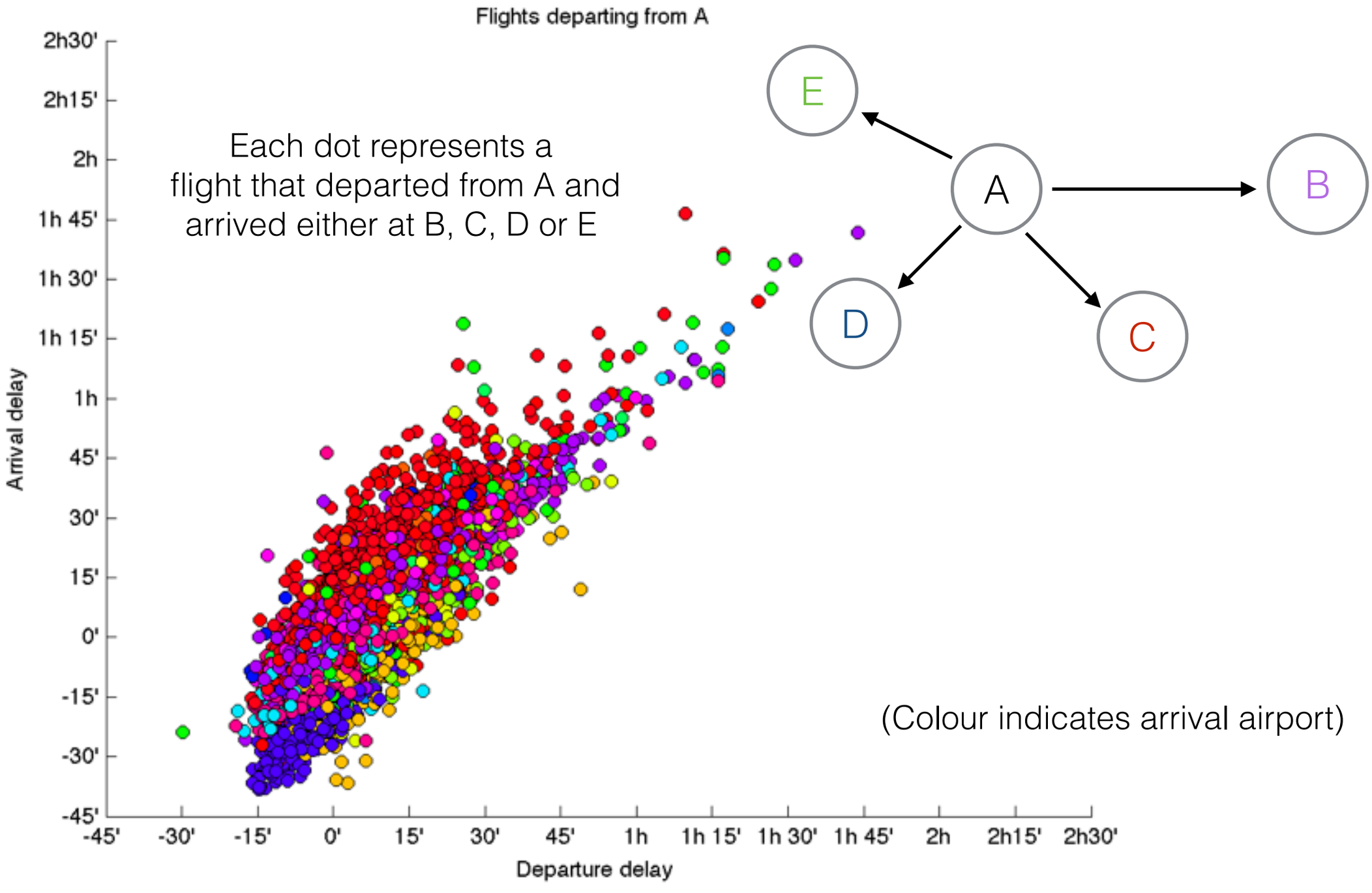


route analysis

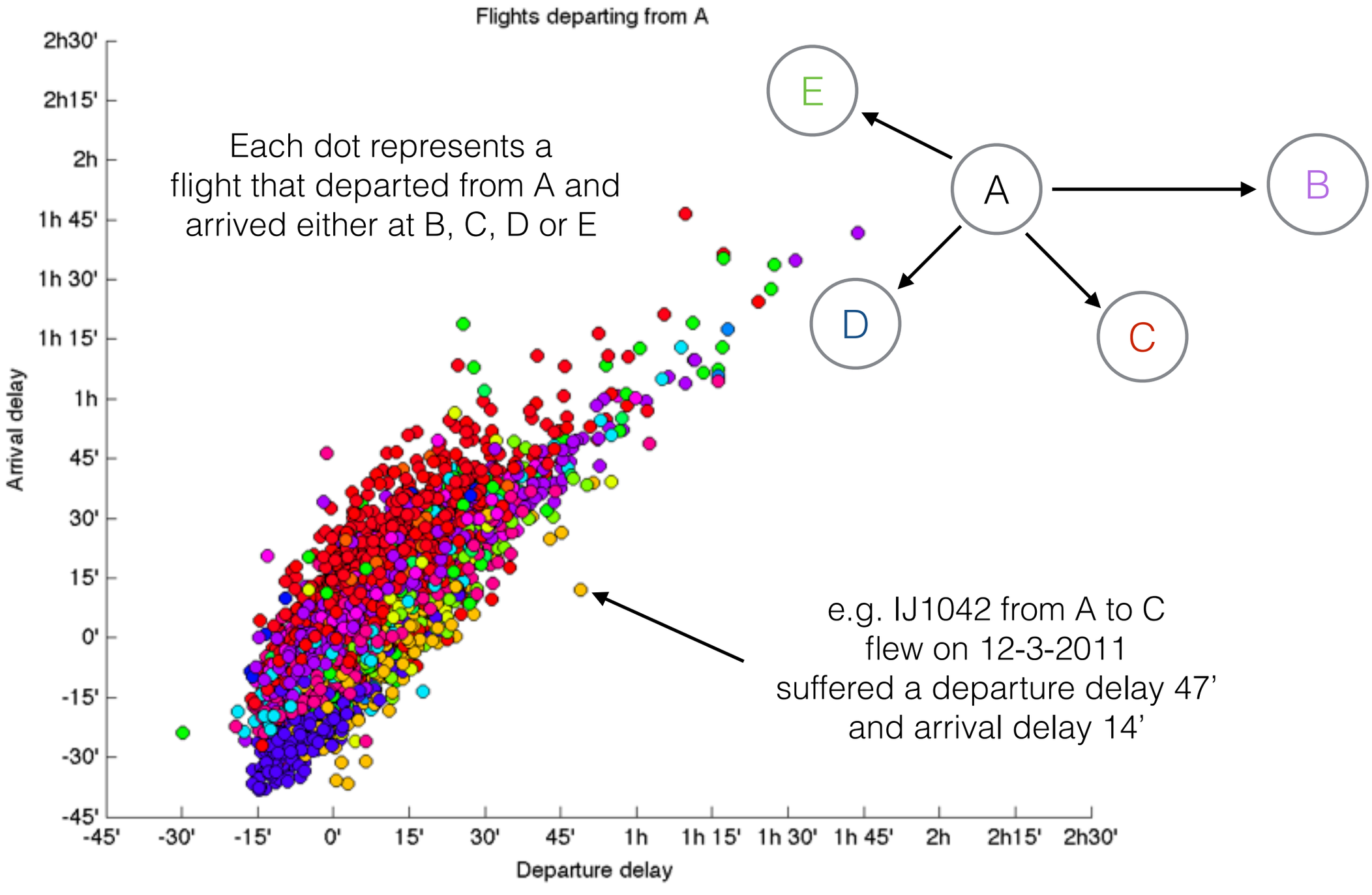


Let us consider all flights departing
from a given airport A. Each flight has a
departure and arrival delay (or zero if on-time).

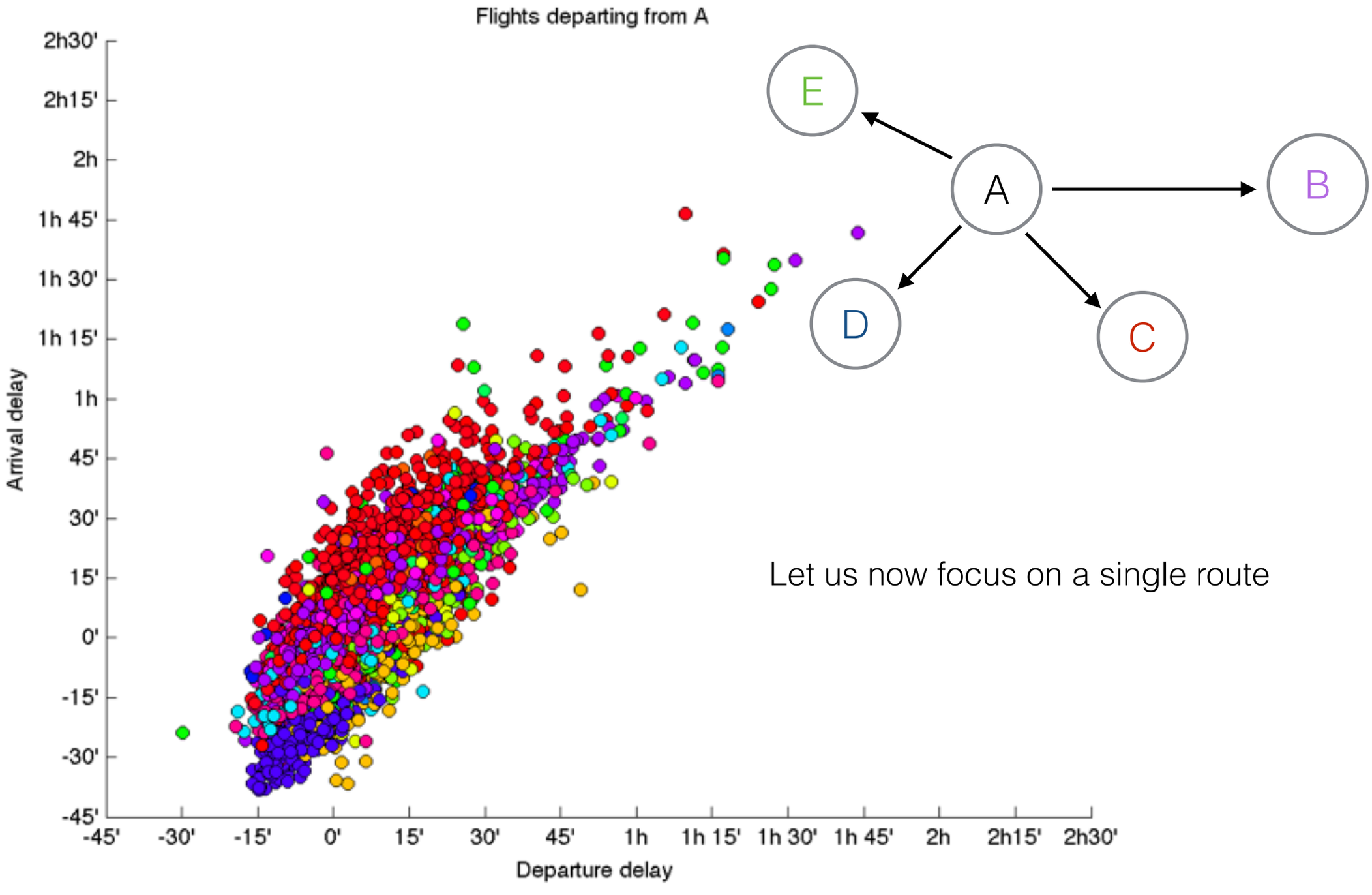
departure vs arrival delay



departure vs arrival delay

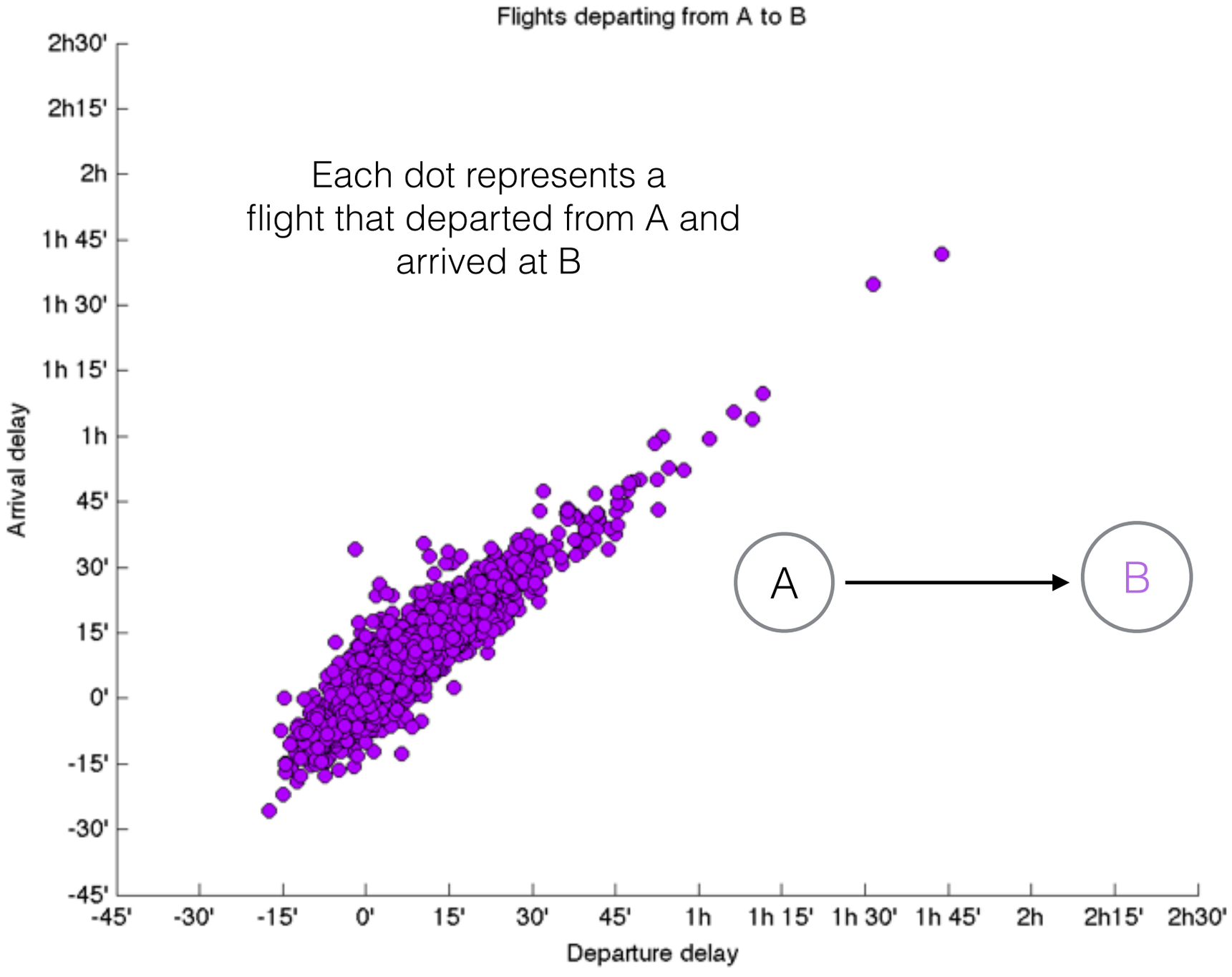


departure vs arrival delay



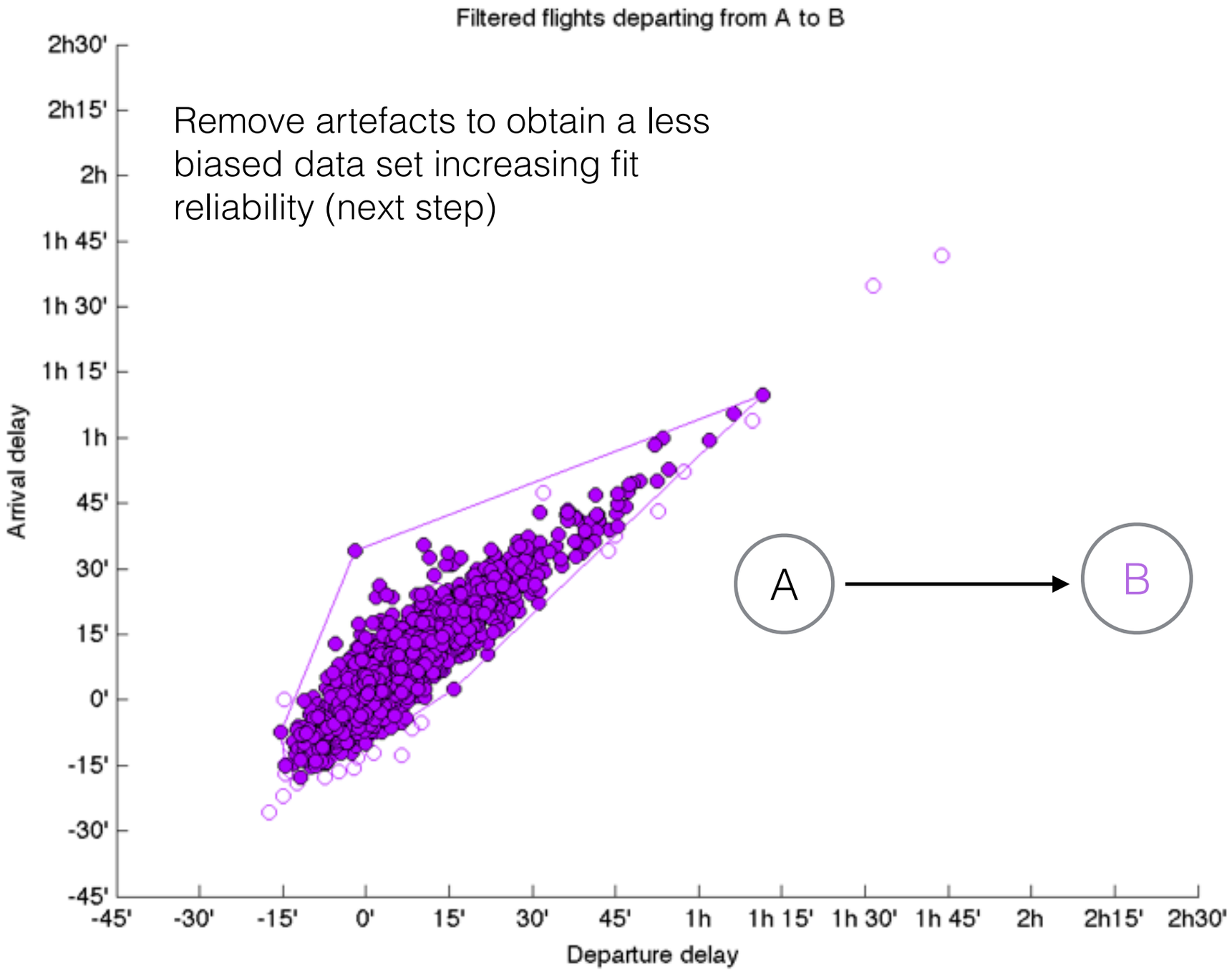
Let us now focus on a single route

route selection

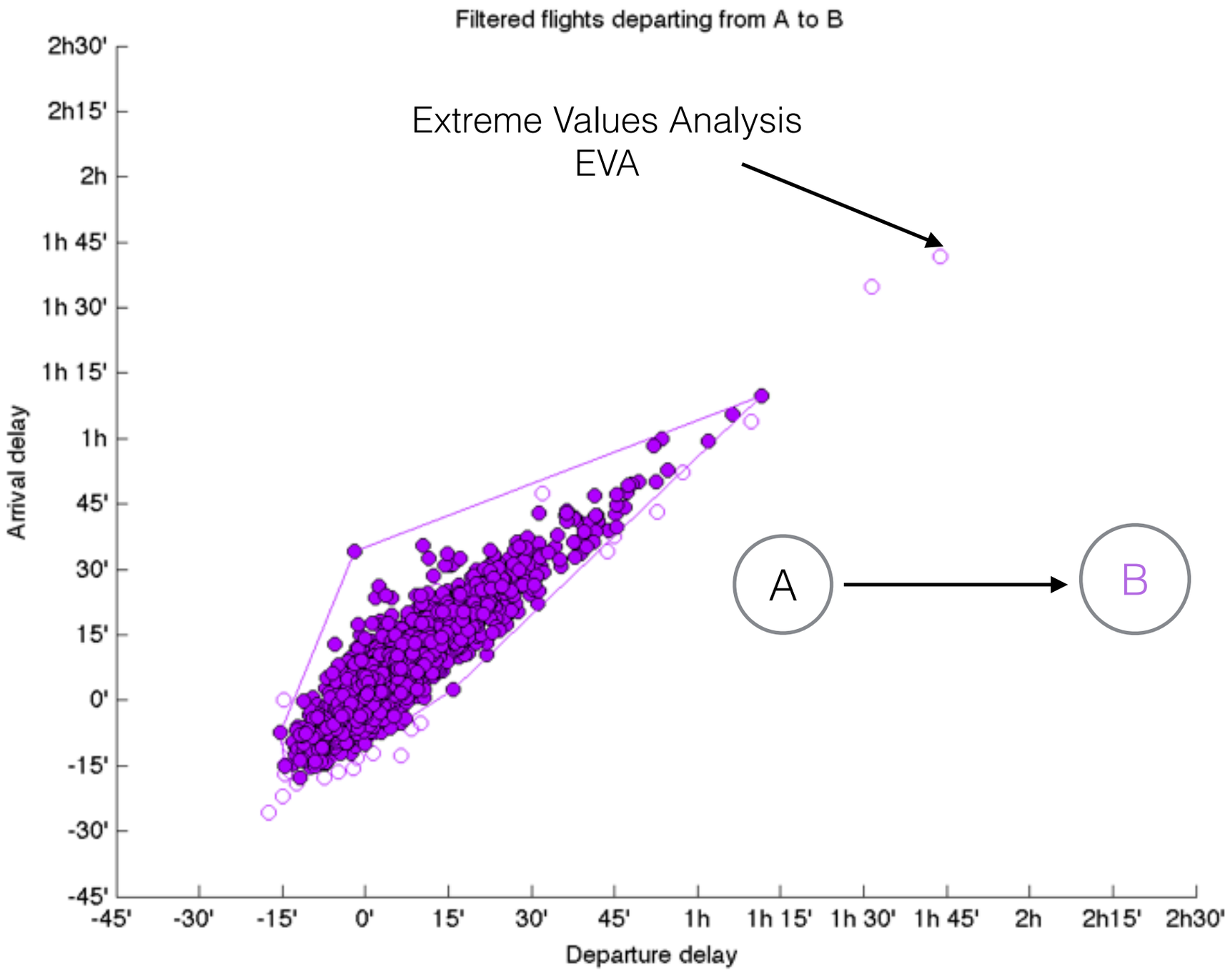


data cleanse

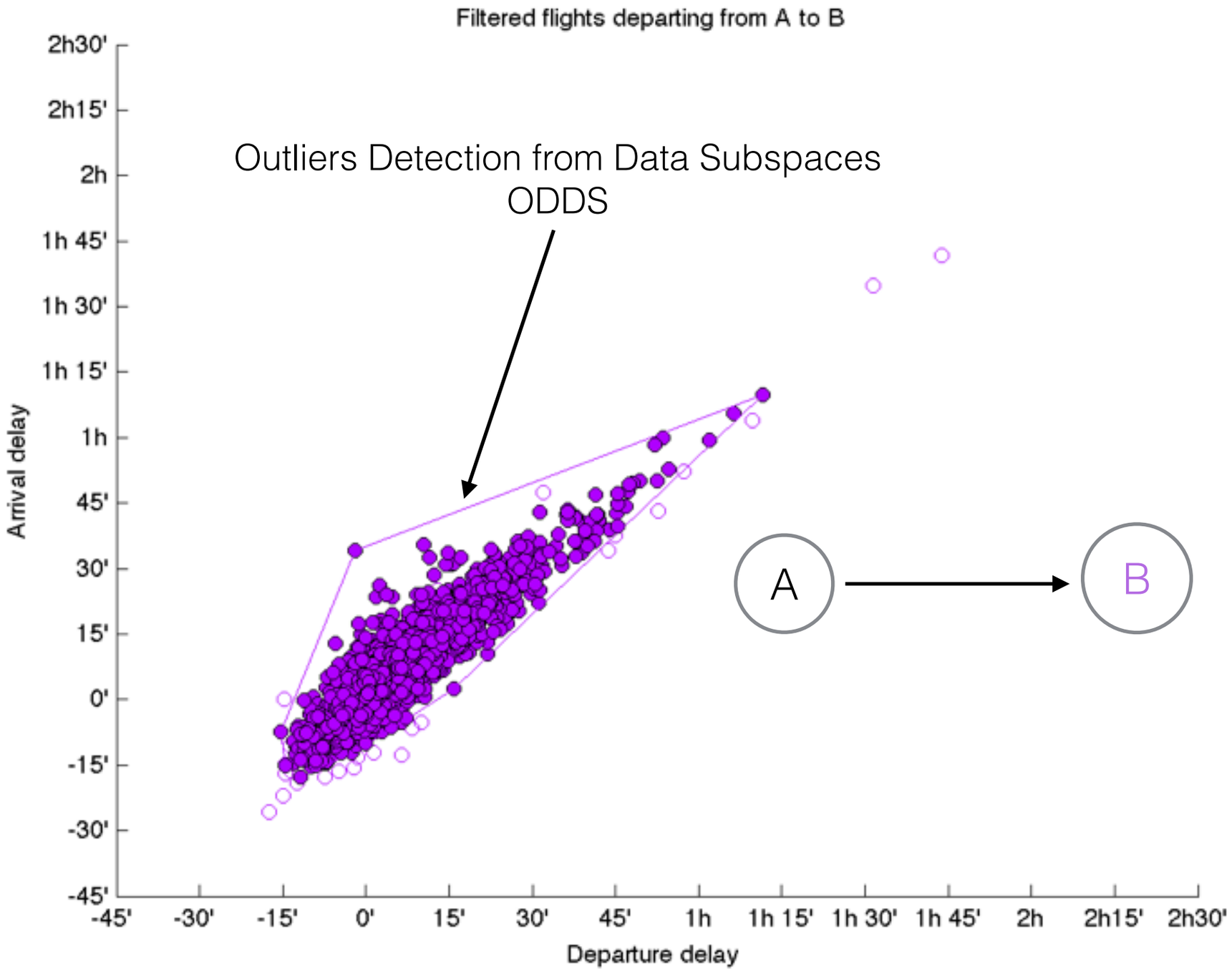
Remove artefacts to obtain a less biased data set increasing fit reliability (next step)



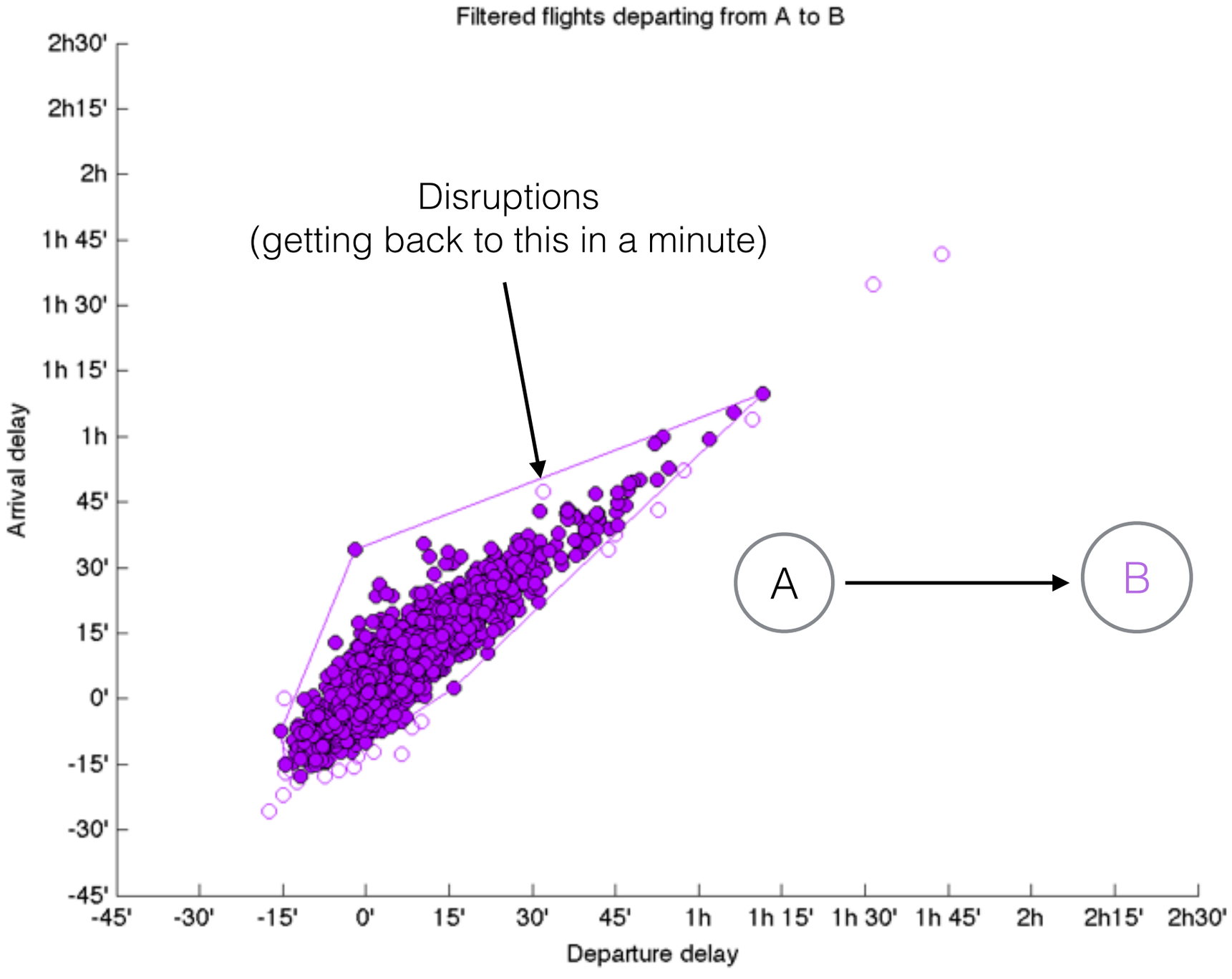
data cleanse



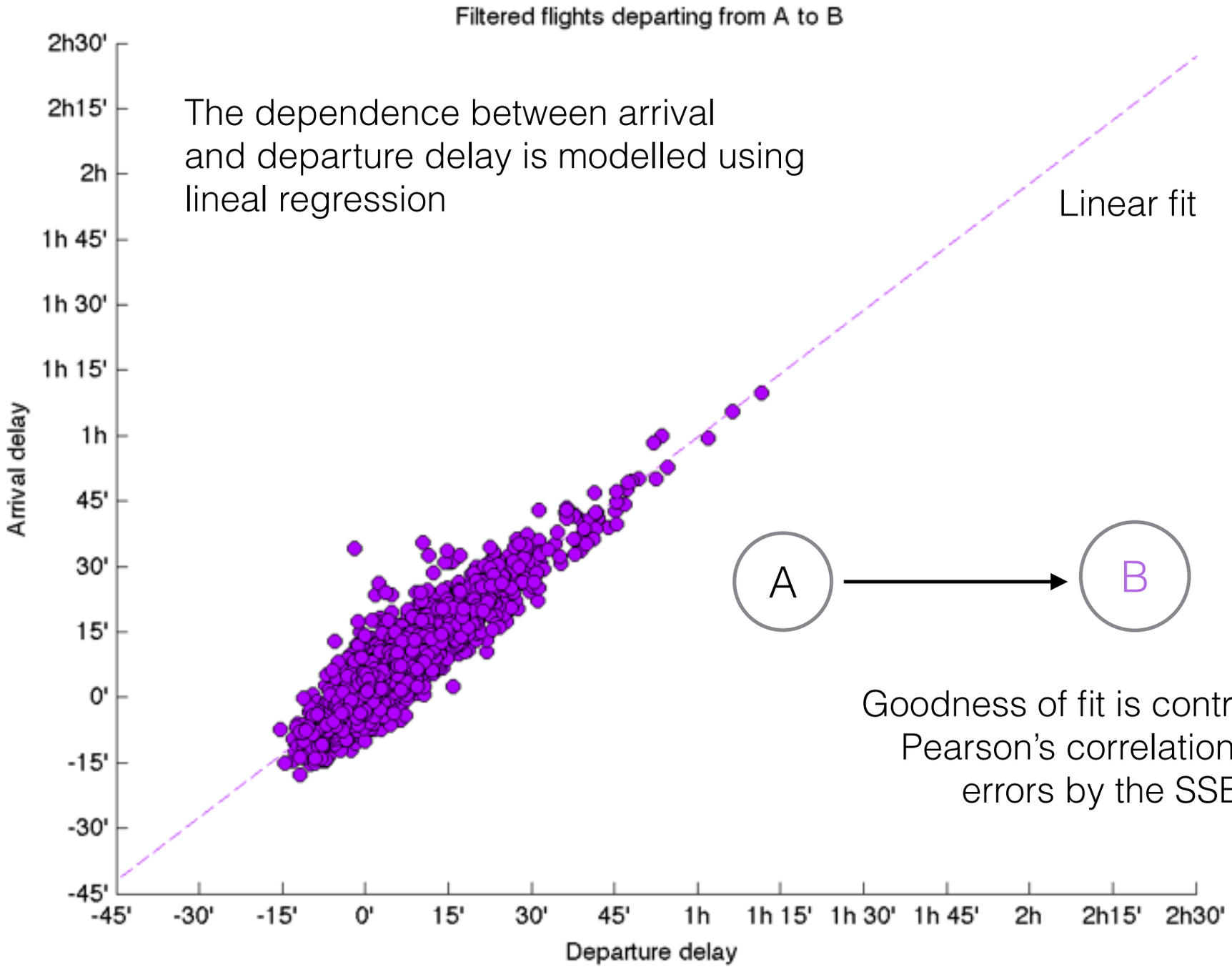
data cleanse



data cleanse



data fitting

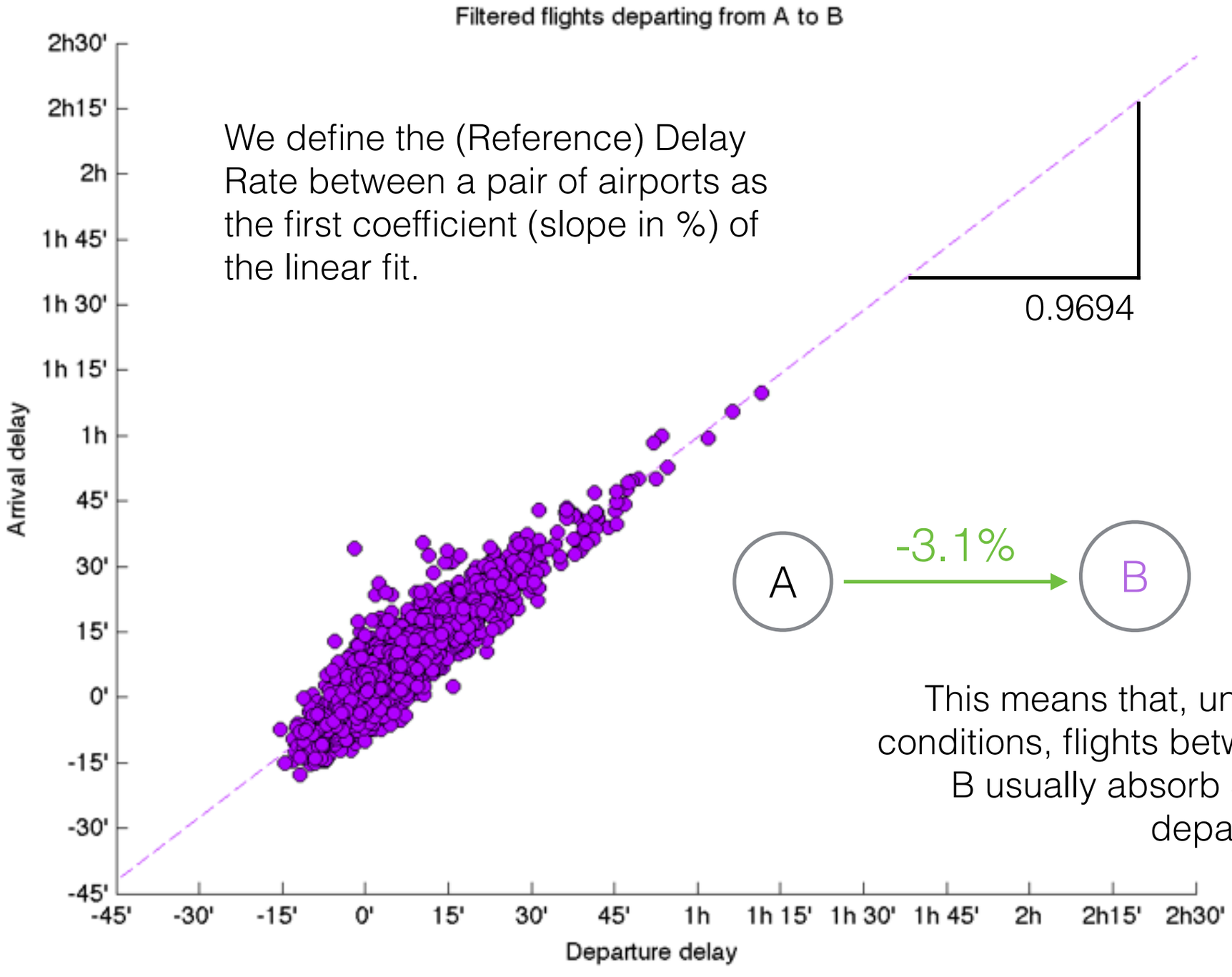


The dependence between arrival and departure delay is modelled using lineal regression

Linear fit

Goodness of fit is controlled by the Pearson's correlation coefficient, errors by the SSE and RSME values.

reference delay rate (RDR)



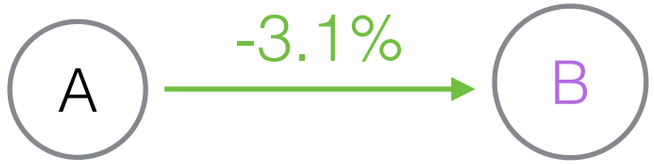
We define the (Reference) Delay Rate between a pair of airports as the first coefficient (slope in %) of the linear fit.

0.9694

A -3.1% B

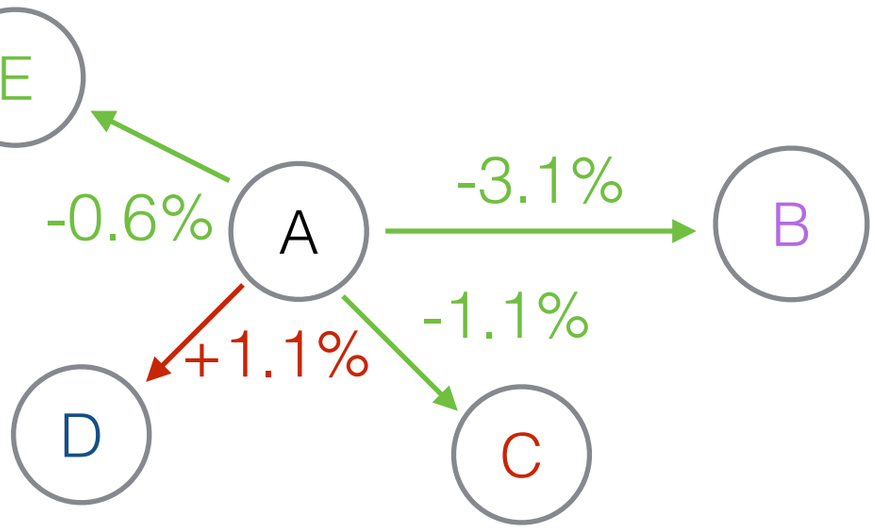
This means that, under normal conditions, flights between A and B usually absorb 3.1% of the departure delay.

reference delay rate (RDR)



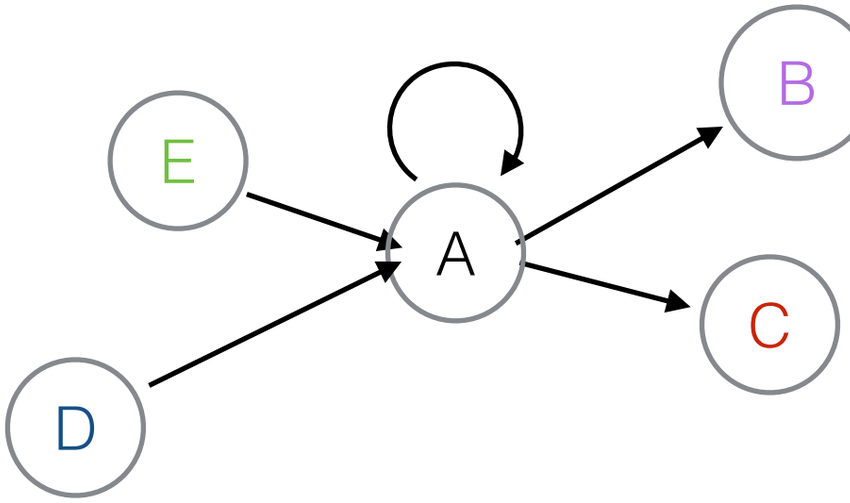
A positive RDR implies delay amplification whilst a negative value implies delay absorption

reference delay rate (RDR)



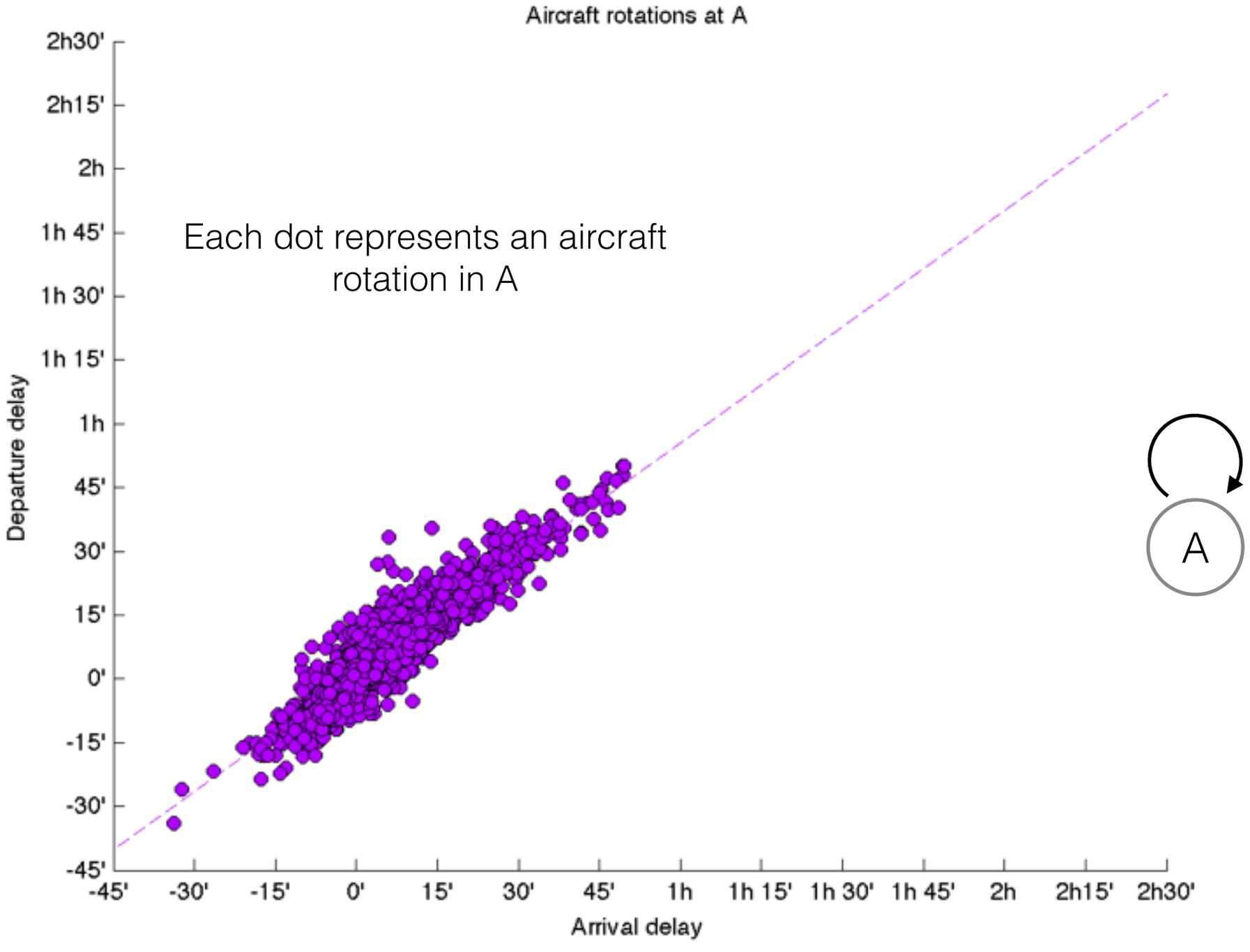
The whole network of delay amplification/absorption can be created using the same technique over all routes, but a very important source of delay is missing: **turnaround** delay.

turnaround delay

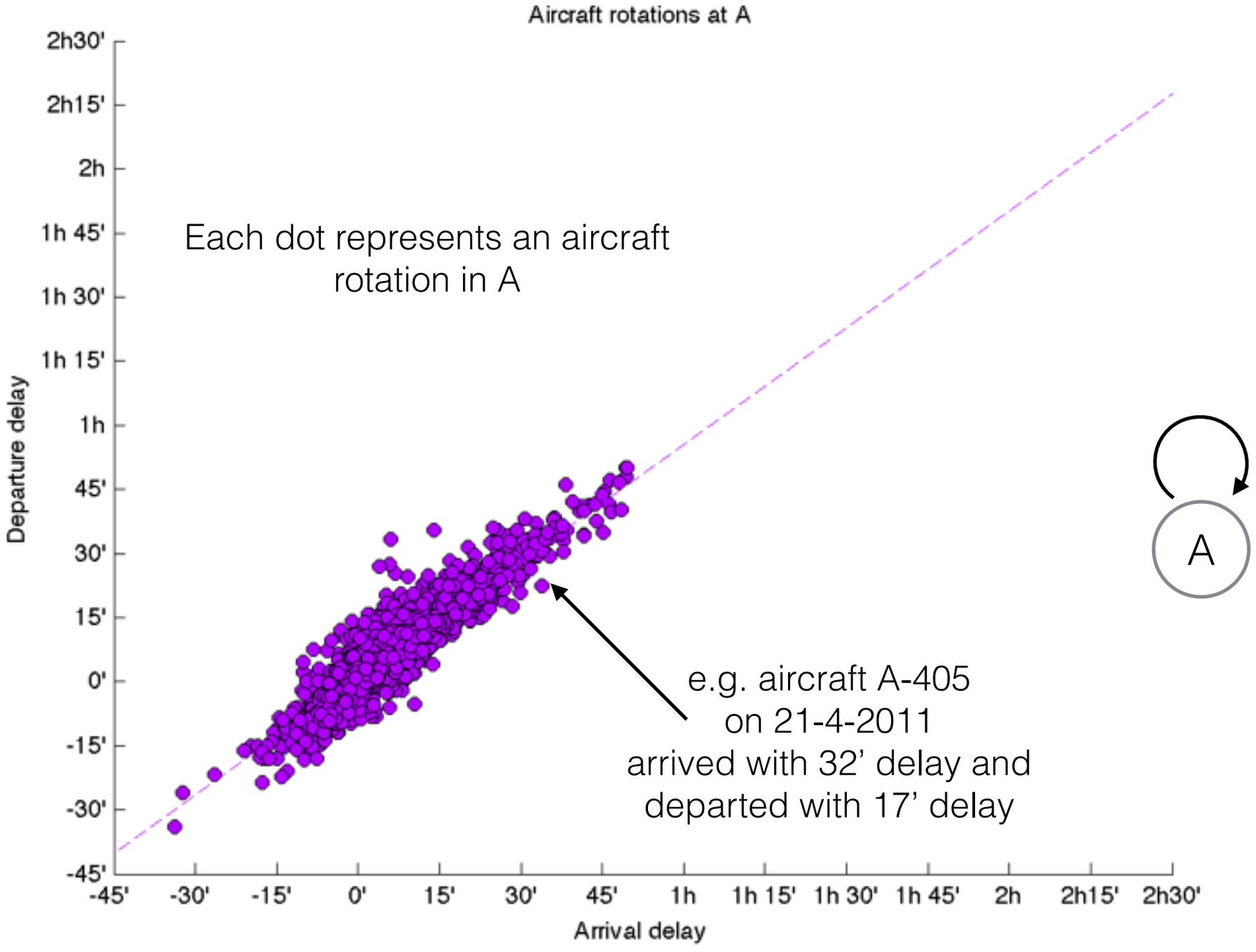


Let us consider all aircraft rotations at an airport A.
Each rotation would have an associated **arrival delay of the previous leg and a departure delay of the next leg** (or zero if on-time).

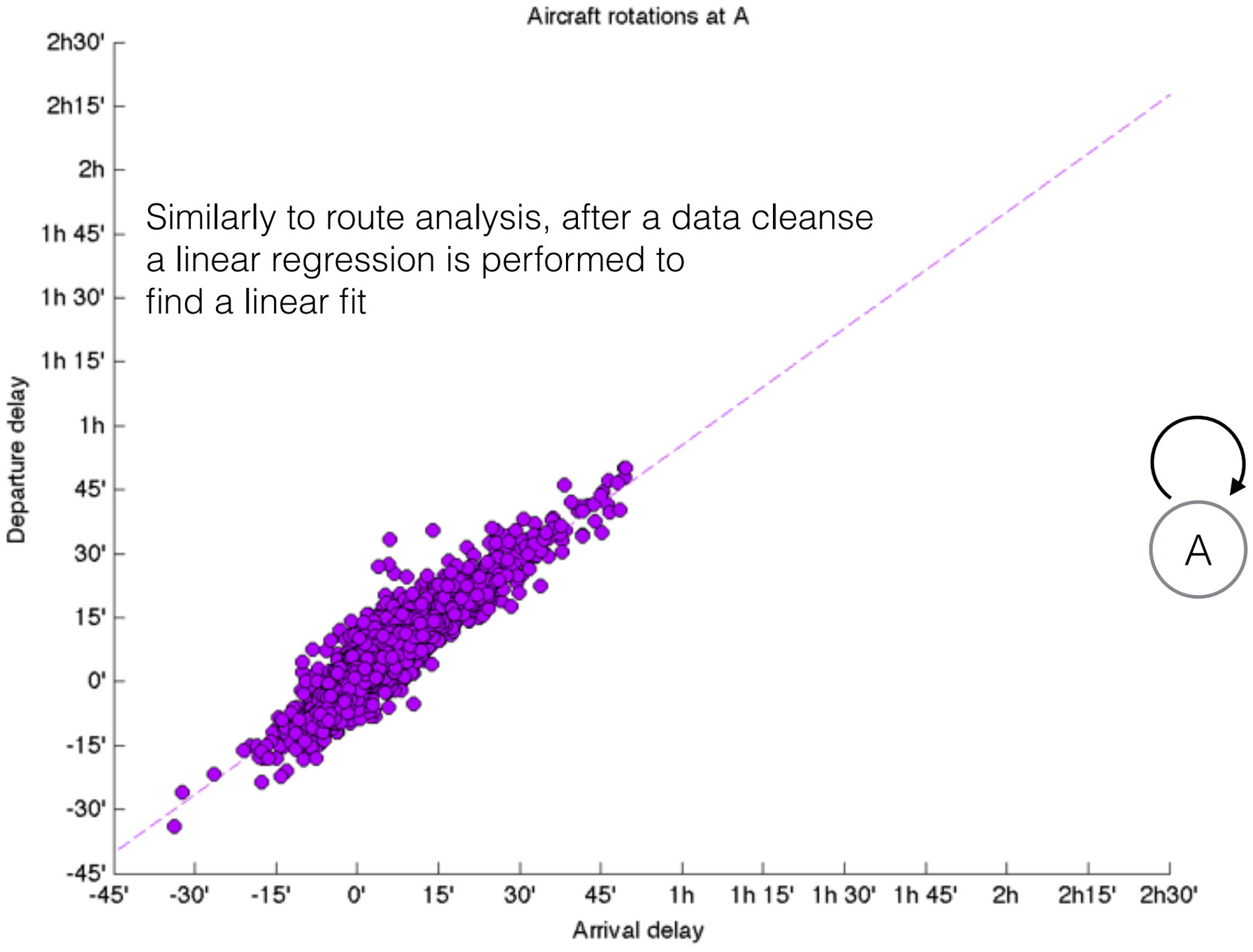
turnaround delay



turnaround delay

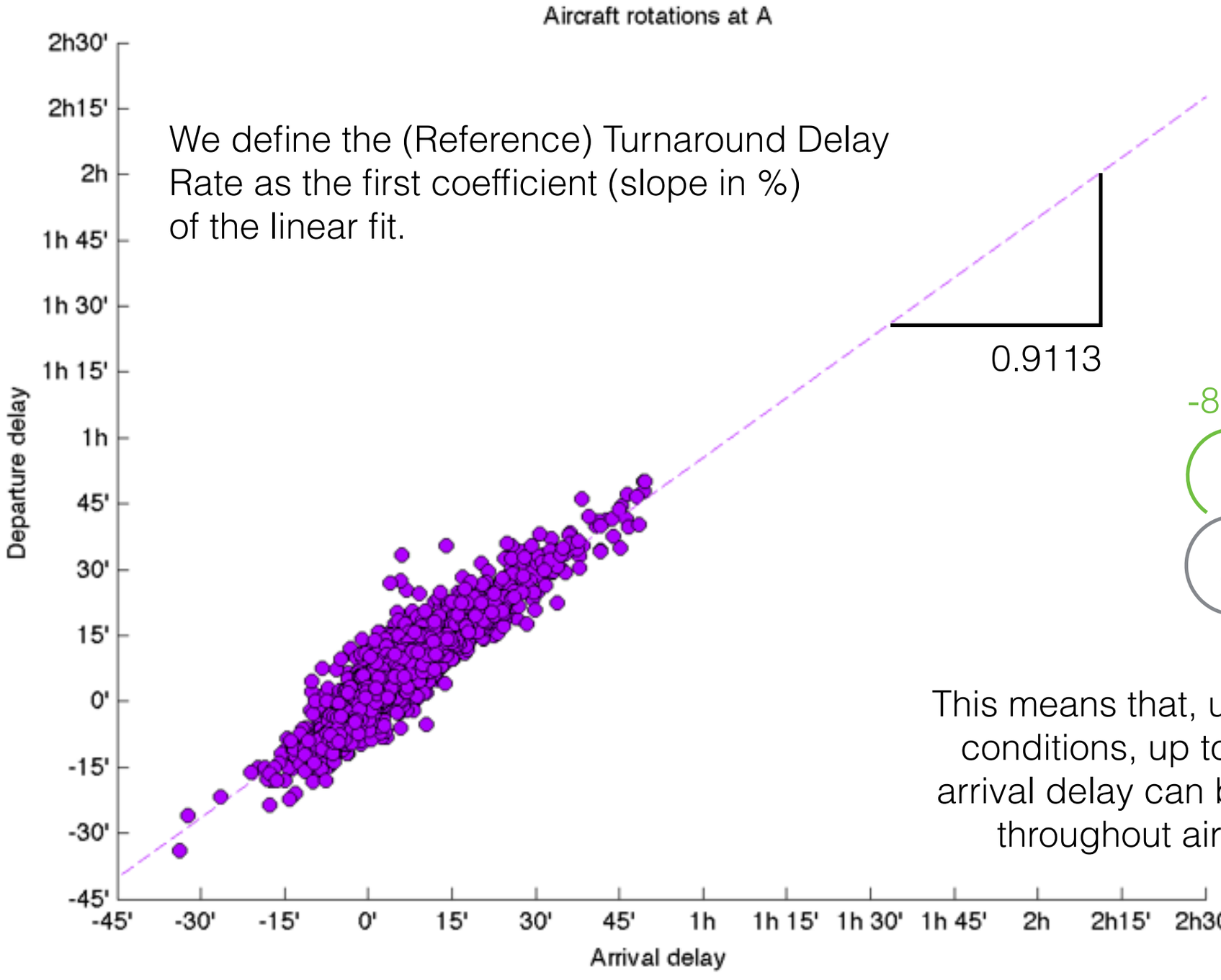


turnaround delay



Similarly to route analysis, after a data cleanse
a linear regression is performed to
find a linear fit

turnaround delay



We define the (Reference) Turnaround Delay Rate as the first coefficient (slope in %) of the linear fit.

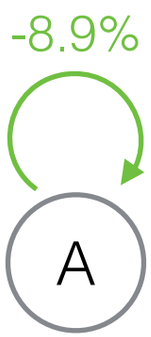
0.9113

-8.9%

A

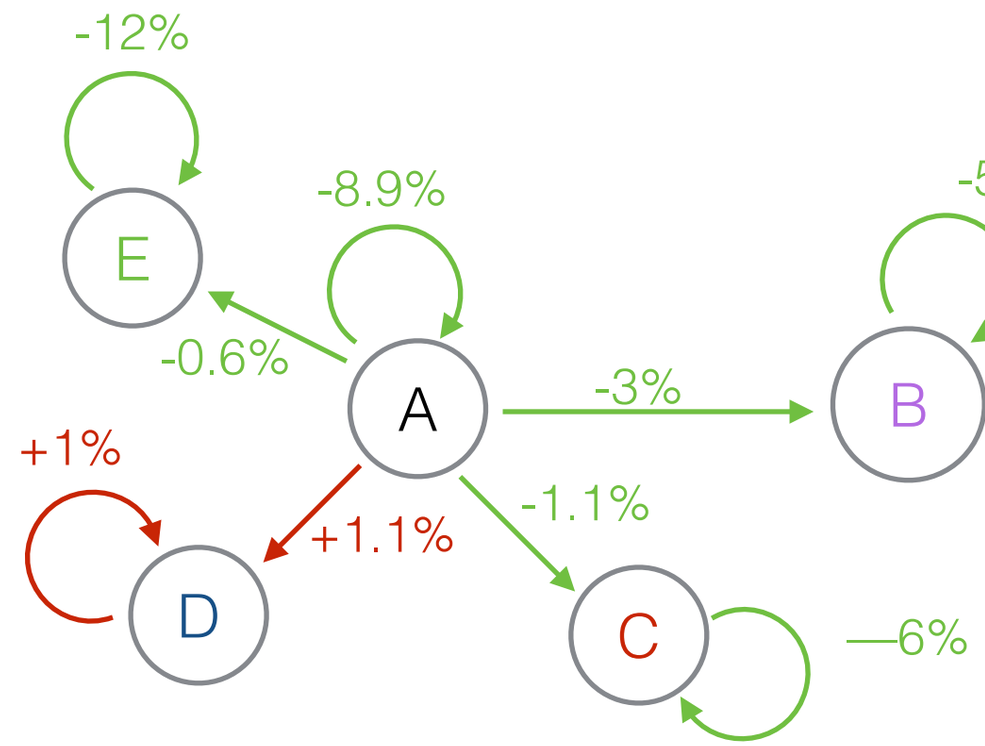
This means that, under normal conditions, up to 8.9% of the arrival delay can be absorbed throughout aircraft rotation

turnaround delay



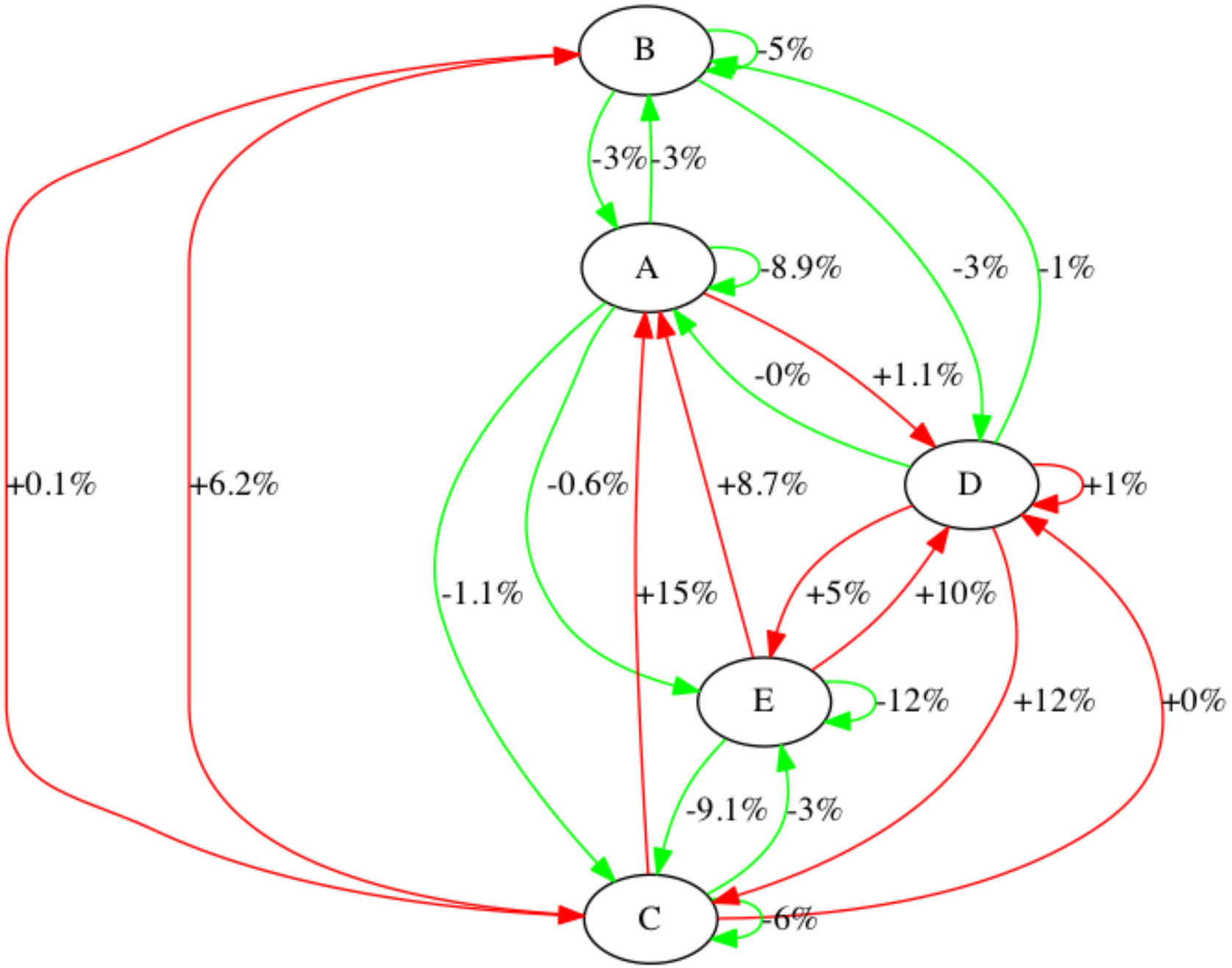
A positive RTDR implies delay amplification at airport A whilst a negative value implies delay absorption

the reference Graph



The process repeats for each airport and it is combined with the route analysis producing a graph picture of (reference) delay rates

the reference Graph



resilience again

- Going back to the resilience concept:

  Resilience is the ability to recover under abnormal conditions.

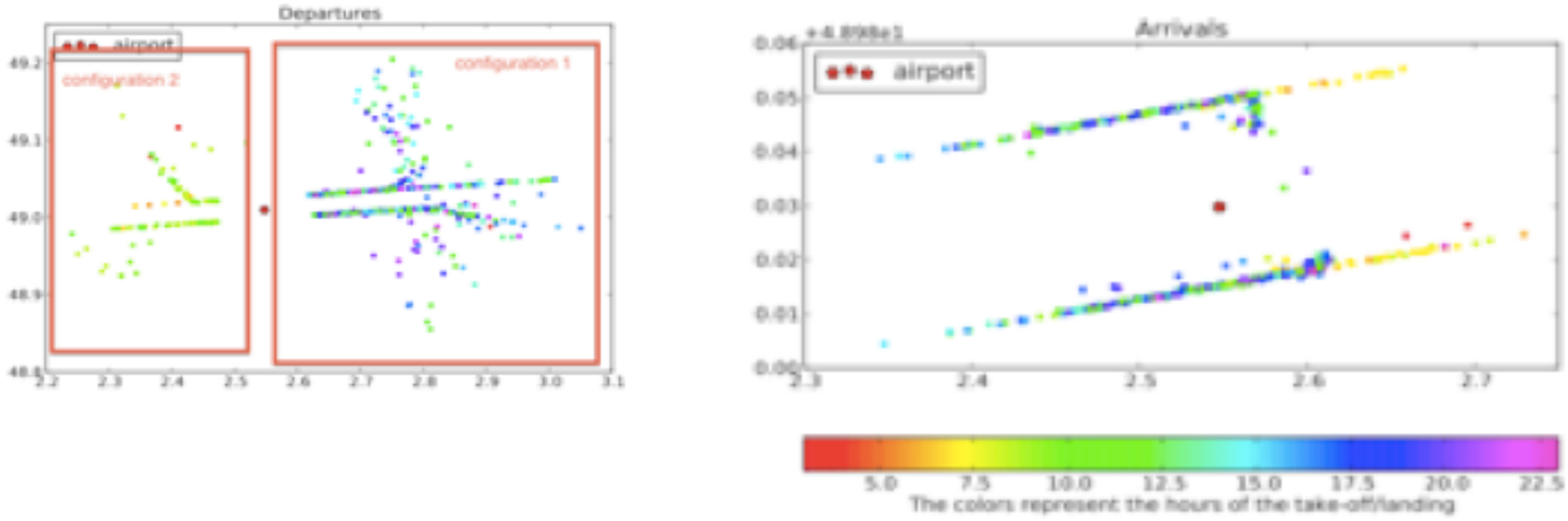
- What would be the best way to measure the performance degradation due to abnormal conditions?

- What are those abnormal conditions? Disruptions!!

- To find out which flights are affected by a disruption external sources of information need to be explored.

disruption detection

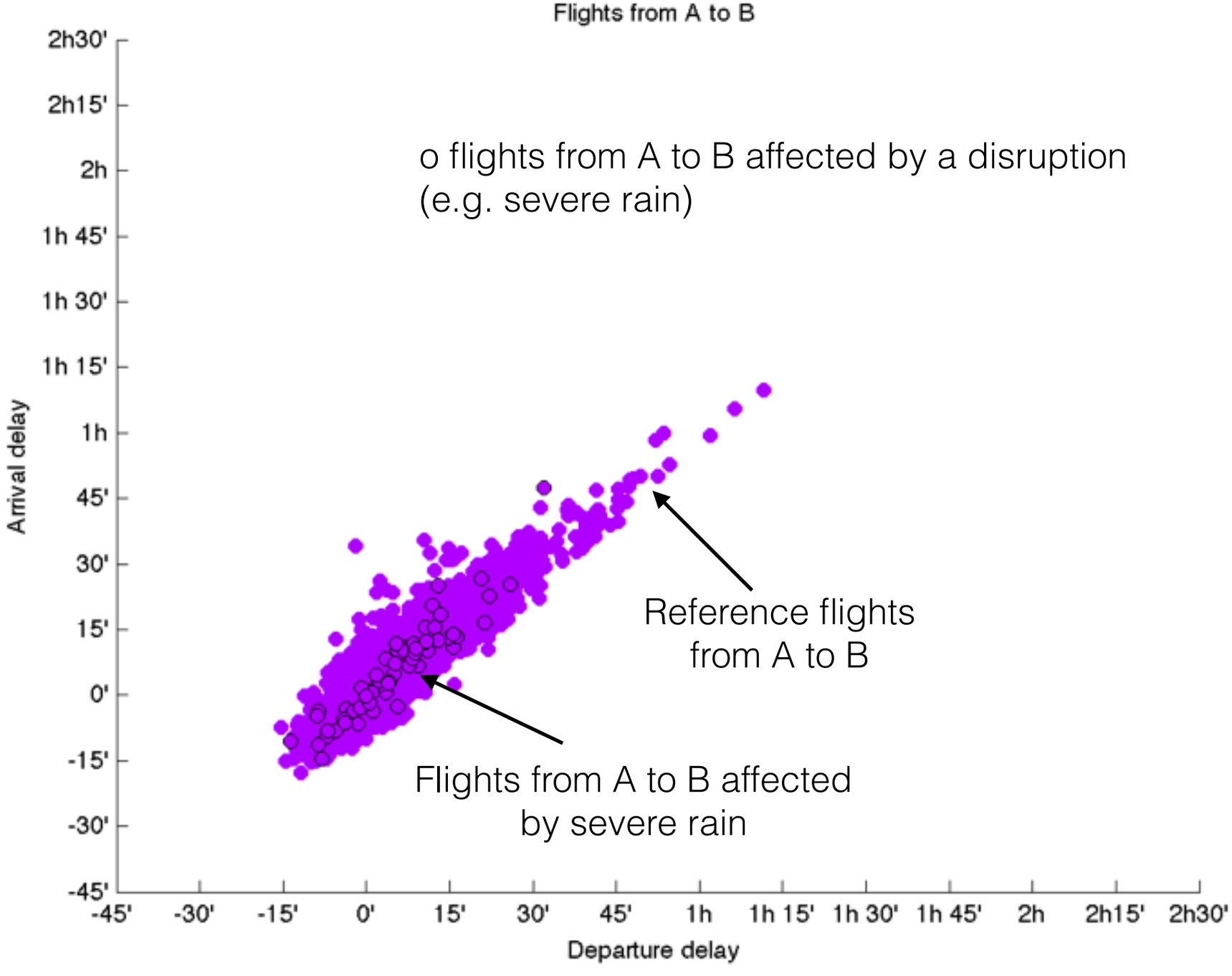
- Runway configuration change by first radar point data clustering:



- Events: ACC headlines' sentences processing

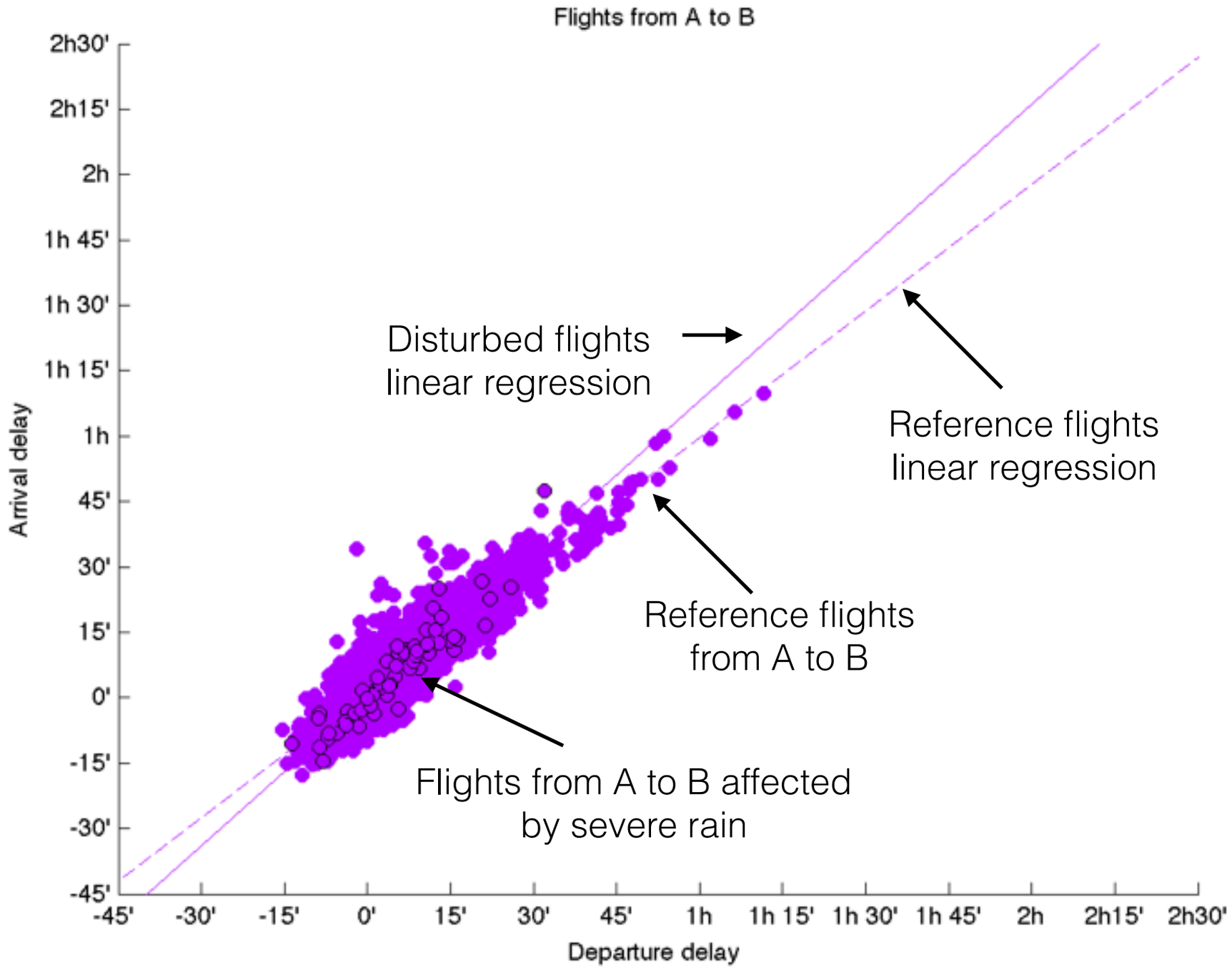
9-Dec-11,"EDDM, EHAM, LPPT, LPPR, EGLL, EHAM weather regulations. EFHK regulations due to snow and strong winds. GCLP closed due to military parade and fly past. LIRP delays due to VOR/DME testing. ENGM regulations due to staffing. LFFF ODIL, TE, OKIP regulations due to Grenelle procedure. EPWW regulations due to capacity in D sector and staffing in R sector. GCCC regulations in INB sector due to capacity."

delay under disruption

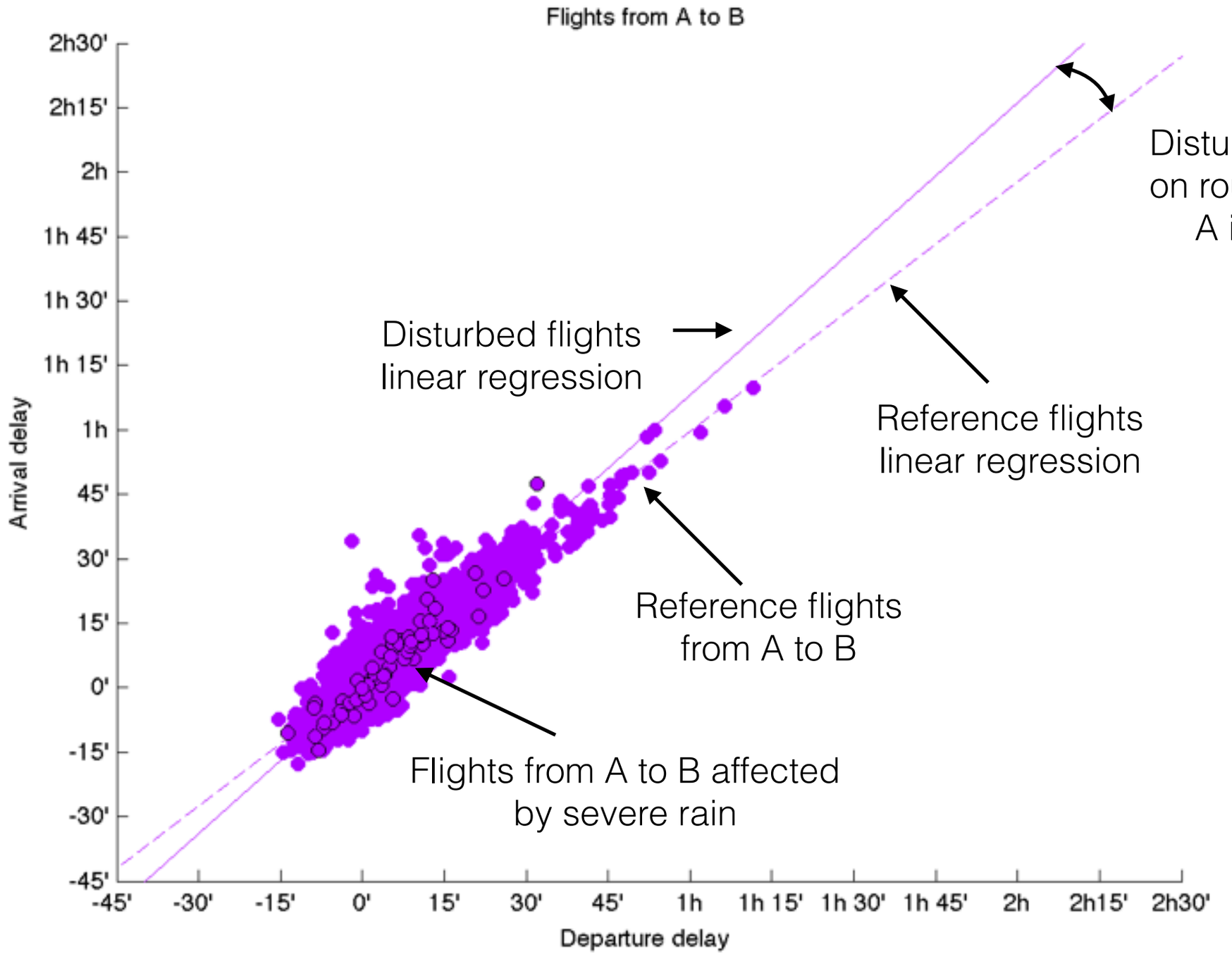


o flights from A to B affected by a disruption (e.g. severe rain)

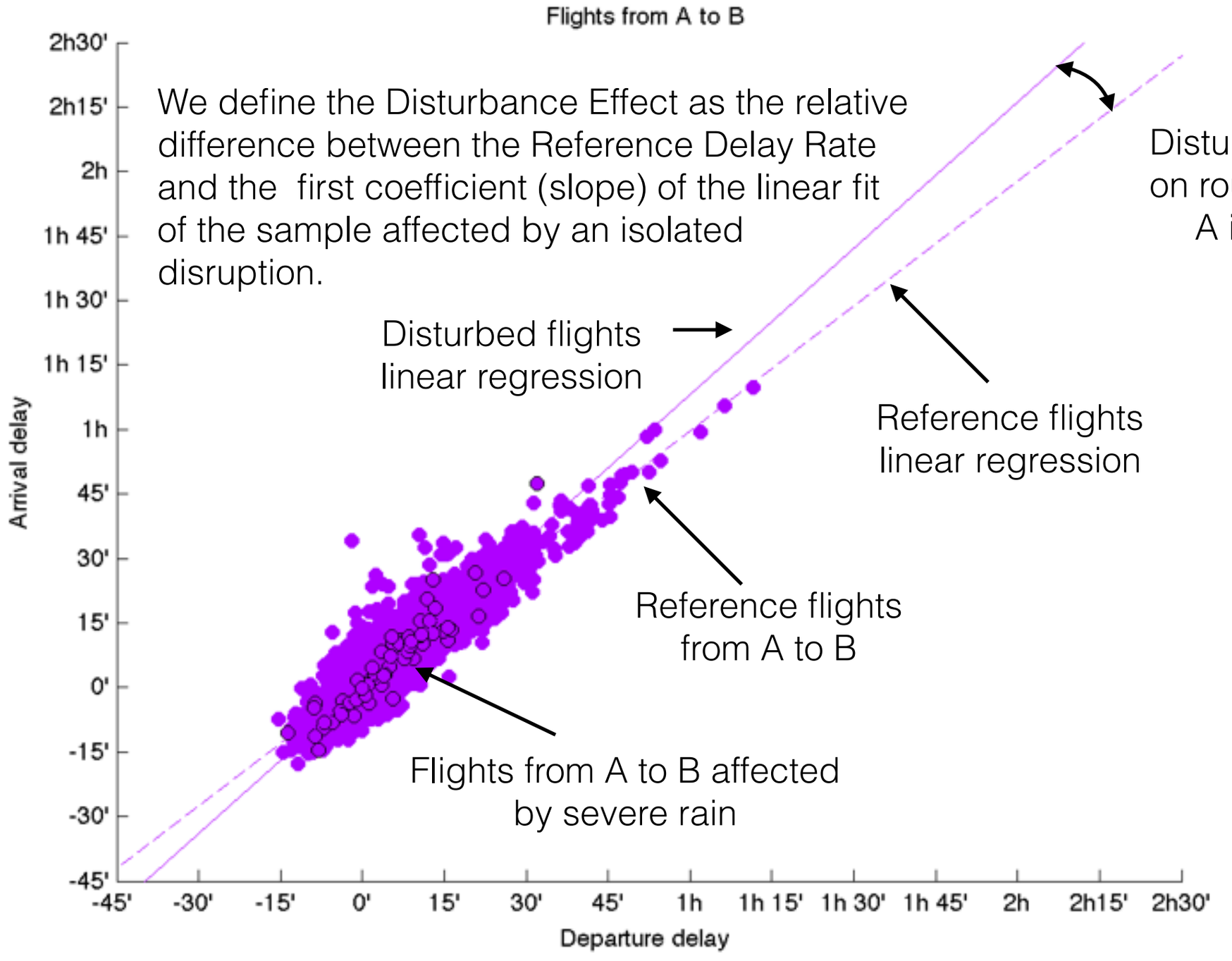
delay under disruption



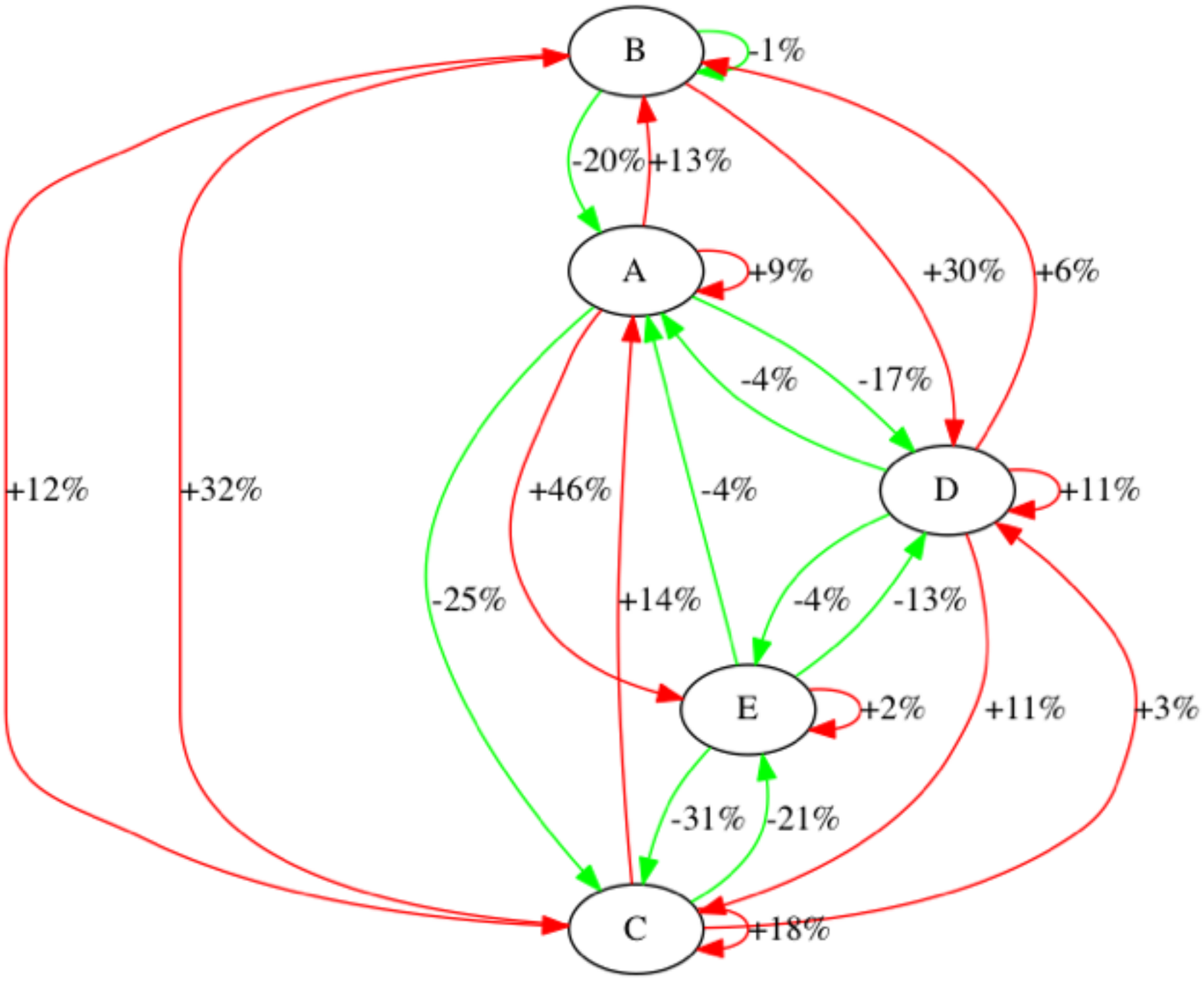
delay under disruption



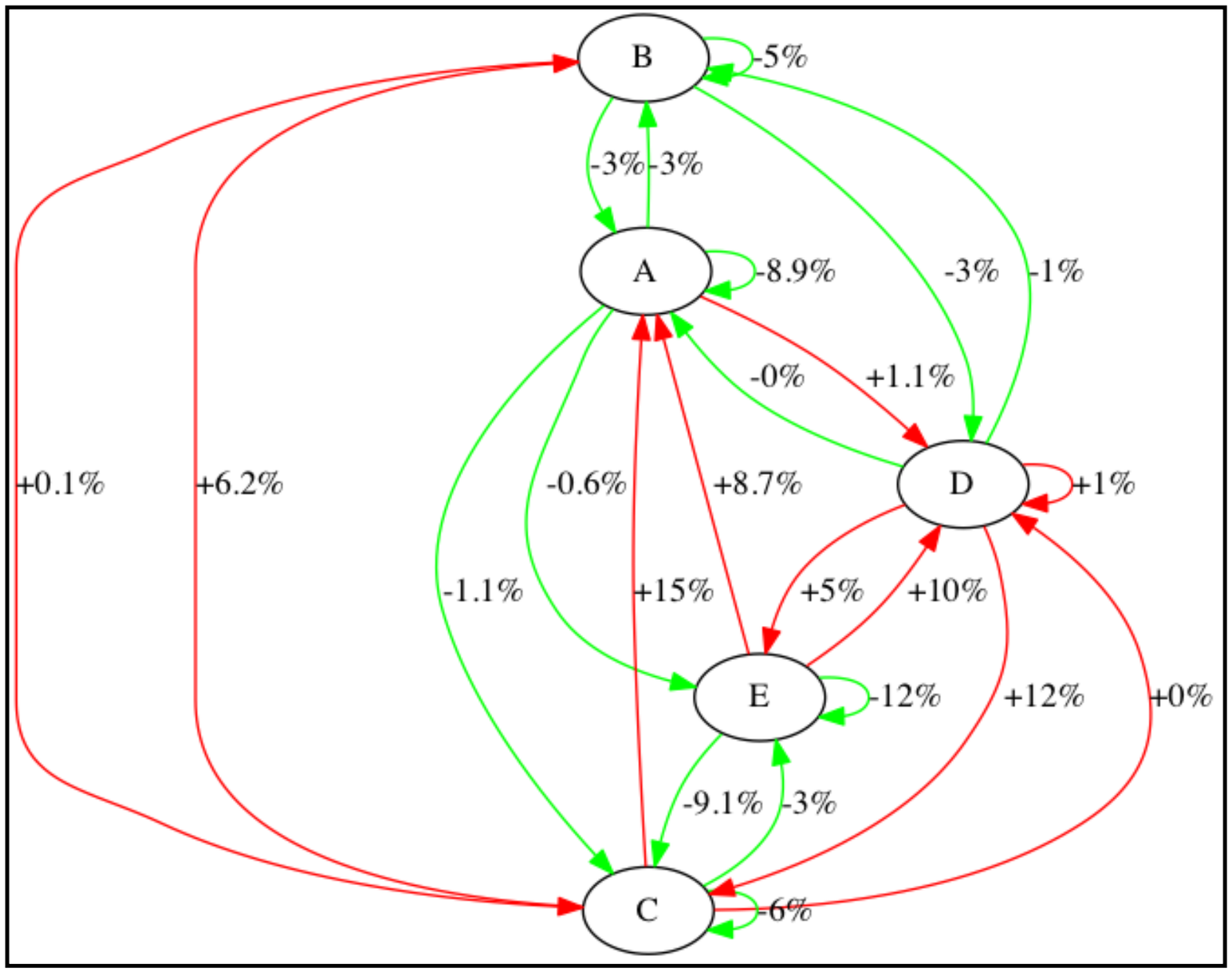
delay under disruption



disturbance Graph

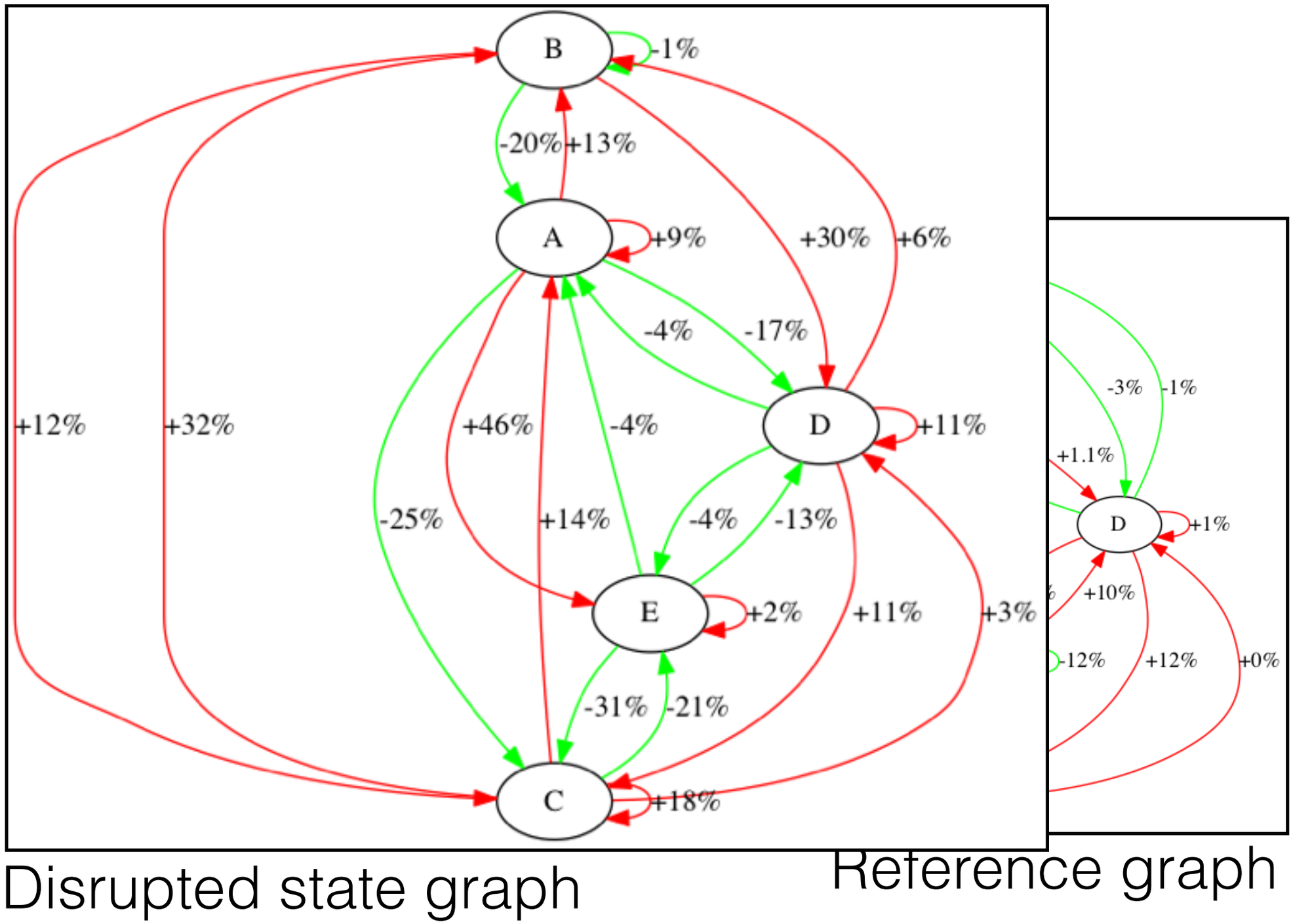


multilayered Graph



Reference graph

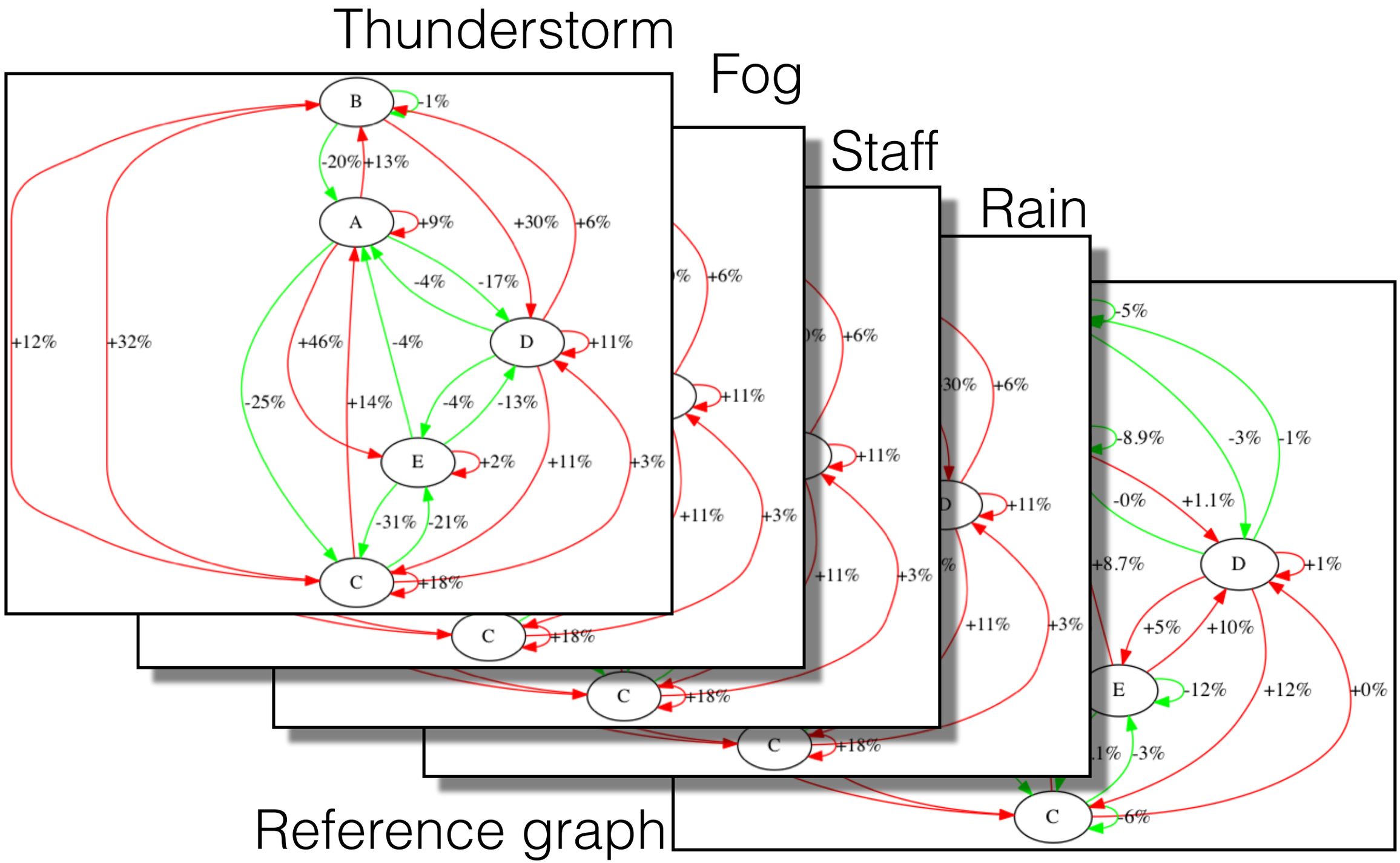
multilayered Graph



Disrupted state graph

Reference graph

multilayered Graph



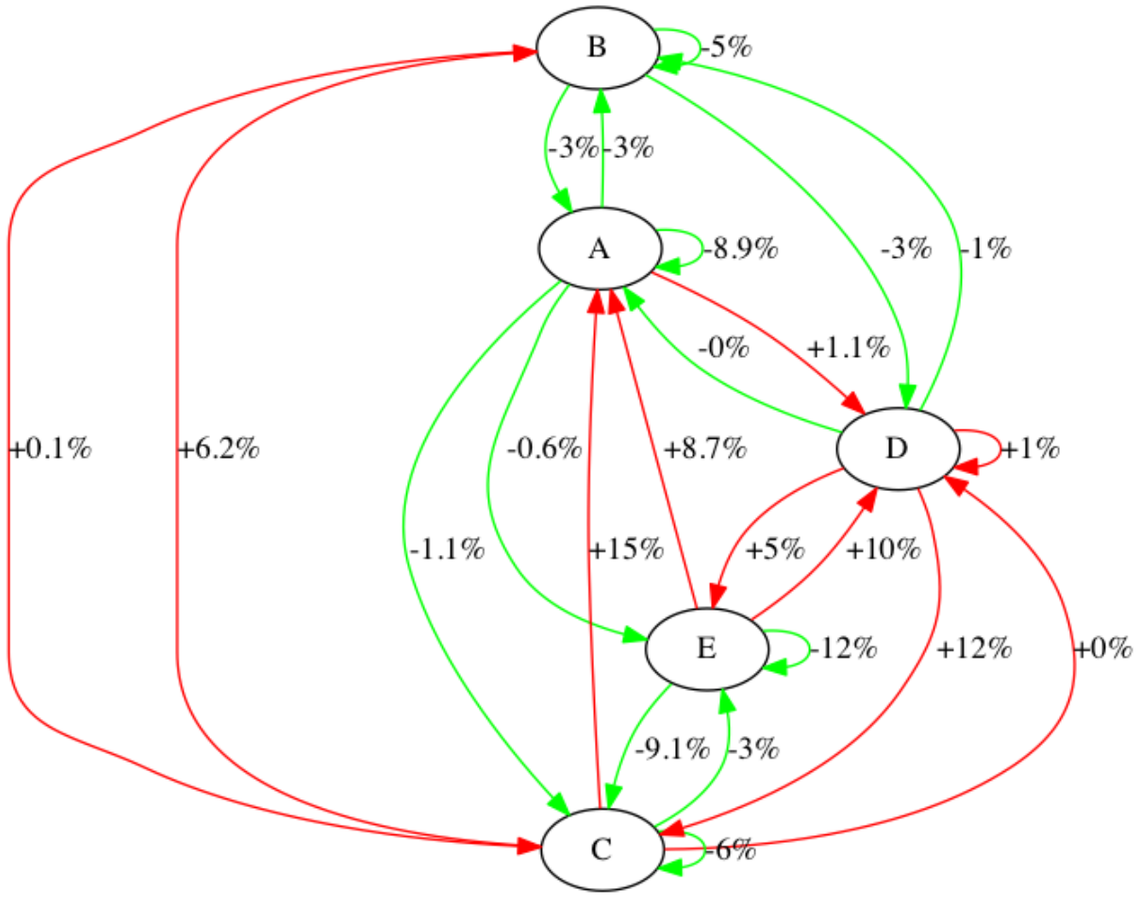
Reference graph

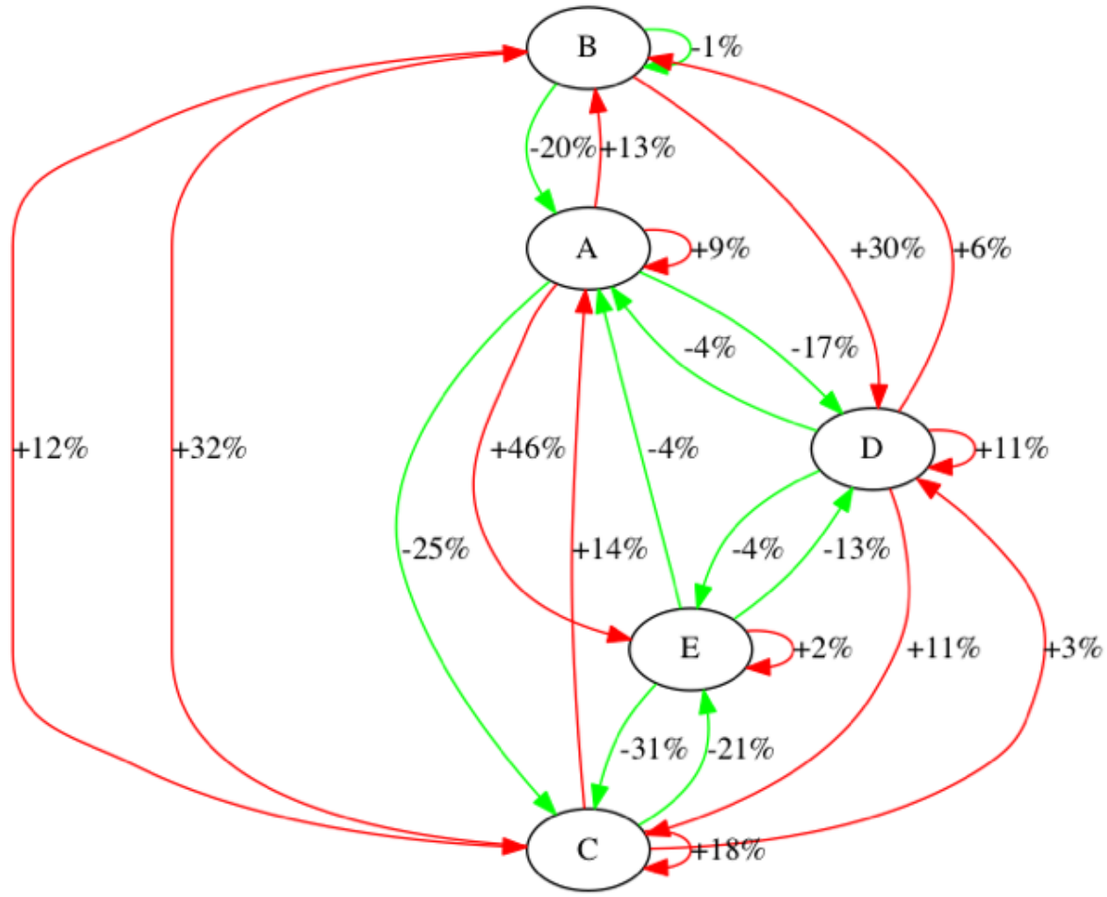
data Visualisation

```
DLH7MA,DAILE,EDDM,EBBR,visibility,500.0,454.0,2011-11-03 09:03:34
DLH2WU,DAIZB,EDDM,EGLL,visibility,500.0,742.0,2011-11-03 09:13:22
DLH9MJ,DAIQN,EDDM,EDDF,visibility,300.0,141.0,2011-11-06 05:03:21
CFG096,DAICG,EDDM,LEPA,visibility,300.0,468.0,2011-11-06 05:09:48
BER9822,DABAG,EDDM,LEPA,visibility,300.0,6.0,2011-11-06 05:12:06
BER56C,DABKP,EDDM,LEPA,visibility,300.0,42.0,2011-11-06 05:12:42
SIA328,9VSWT,EDDM,EGCC,visibility,300.0,-169.0,2011-11-06 05:18:11
DLH9MJ,DAISE,EDDM,EDDF,visibility,430.0,472.0,2011-11-08 05:08:52
SRR6118,OYSRI,EDDM,LGAV,visibility,430.0,9.0,2011-11-08 05:09:09
BER56C,DAHFA,EDDM,LEPA,visibility,430.0,30.0,2011-11-08 05:12:30
DLH88N,DAEBC,EDDM,LFPG,visibility,360.0,648.0,2011-11-08 05:53:48
DLH2N,DAILW,EDDM,EDDT,visibility,360.0,-342.0,2011-11-08 05:58:18
DLH2000,DAILM,EDDM,EDDL,visibility,360.0,1402.0,2011-11-08 06:27:22
DLH4NC,DACKL,EDDM,EBBR,visibility,360.0,358.0,2011-11-08 06:24:58
DLH101,DAIRR,EDDM,EDDF,visibility,360.0,610.0,2011-11-08 06:29:10
BAW947L,GEUPA,EDDM,EGLL,visibility,360.0,529.0,2011-11-08 06:33:49
JKK134,ECIPI,EDDM,LEBL,visibility,360.0,-975.0,2011-11-08 06:12:45
BER167,DABBG,EDDM,EDDL,visibility,360.0,1631.0,2011-11-08 07:02:11
DLH5T,DAILY,EDDM,EDDT,visibility,360.0,-19.0,2011-11-08 06:40:41
BER6192,DABCA,EDDM,EDDT,visibility,360.0,-102.0,2011-11-08 06:43:18
AFR1123,None,EDDM,LFPG,visibility,360.0,231.0,2011-11-08 06:48:51
DLH2JT,DAISI,EDDM,EDDL,visibility,360.0,904.0,2011-11-08 07:06:04
DLH6HT,DAIPK,EDDM,EGLL,visibility,360.0,-615.0,2011-11-08 06:45:45
IBE35YE,None,EDDM,LEMD,visibility,360.0,-497.0,2011-11-08 07:08:43
DLH6N,DAIZB,EDDM,EDDT,visibility,360.0,-125.0,2011-11-08 07:20:55
DLH7EW,DAIRT,EDDM,EDDF,visibility,360.0,-178.0,2011-11-08 07:30:02
BER630,DABAF,EDDM,EDDL,visibility,360.0,1646.0,2011-11-08 07:59:26
```

```
BAW457,GEUUP,LEMD,EGLL,noDisturbance,None,2276.0,2011-04-26 13:15:28
IBE35LZ,None,LEMD,EDDM,noDisturbance,None,923.0,2011-04-26 20:17:57
ANE8054,ECLJR,LEMD,EDDL,noDisturbance,None,-259.0,2011-04-26 20:51:09
VLG9822,ECKCU,LEMD,LEMG,noDisturbance,None,566.0,2011-04-26 10:59:12
TAP713,CSTTF,LEMD,LPPT,noDisturbance,None,626.0,2011-04-26 11:50:22
DLH17P,DACNR,LEMD,EDDL,noDisturbance,None,632.0,2011-04-26 20:32:25
IBE0216,None,LEMD,LEMG,noDisturbance,None,577.0,2011-04-26 19:08:56
VLG1026,ECLAB,LEMD,LEBL,noDisturbance,None,805.0,2011-04-26 16:15:58
VLG9808,ECLLJ,LEMD,LEPA,noDisturbance,None,-337.0,2011-04-26 16:14:48
DLH12K,DABIK,LEMD,EDDF,noDisturbance,None,-346.0,2011-04-26 11:06:16
IBE19DU,None,LEMD,LEBL,noDisturbance,None,1410.0,2011-04-26 19:24:58
ANGEL50,ECKAE,LEMD,LEBL,noDisturbance,None,-15.0,2011-04-26 03:26:22
JKK446,ECINM,LEMD,LEBL,noDisturbance,None,366.0,2011-04-26 06:46:34
IBE14CN,None,LEMD,LEBL,noDisturbance,None,491.0,2011-04-26 13:55:10
EZY98DH,GEZEB,LEMD,LPPT,noDisturbance,None,-181.0,2011-04-26 19:19:29
VLG1082,ECFQY,LEMD,LEBL,noDisturbance,None,-123.0,2011-04-26 19:58:46
JKK5614,ECGVO,LEMD,LEPA,noDisturbance,None,260.0,2011-04-26 09:28:20
AEA6049,ECISE,LEMD,LEPA,noDisturbance,None,-339.0,2011-04-26 20:27:13
AEA2005,ECISN,LEMD,LEBL,noDisturbance,None,-505.0,2011-04-26 17:43:58
IBE3400,None,LEMD,LFPO,noDisturbance,None,55.0,2011-04-26 20:48:40
BAW455,GBNWZ,LEMD,EGLL,noDisturbance,None,413.0,2011-04-26 07:15:52
IBE3500,None,LEMD,EDDF,noDisturbance,None,1747.0,2011-04-26 09:44:00
IBE0800,None,LEMD,LEBL,noDisturbance,None,566.0,2011-04-26 07:10:58
JKK414,ECKJE,LEMD,LEBL,noDisturbance,None,514.0,2011-04-26 08:29:46
IBE6824,None,LEMD,LEBL,noDisturbance,None,473.0,2011-04-26 08:38:46
ANE8052,ECJNB,LEMD,EDDL,noDisturbance,None,1073.0,2011-04-26 09:44:58
AEA8701,ECISN,LEMD,LEPA,noDisturbance,None,-719.0,2011-04-26 08:21:56
IBE3436,None,LEMD,LEPO,noDisturbance,None,857.0,2011-04-26 09:10:53
```

data Visualisation





the future: dashboards

- Data representation should no longer be static.
- Representation of non-homogeneous data sources.
- Consuming real-time data feeds.
- Interactive and intuitive.
- Extensive use of visual dashboards such as d3.js

Thank you for your attention

innaxis

innovation for
a complex world

Samuel Cristóbal

scristobal@innaxis.org