Queen Mary University of London

Looking for Hypergiants in PeeringDB

Timm Böttger, Felix Cuadrado and Steve Uhlig

timm.boettger@qmul.ac.uk

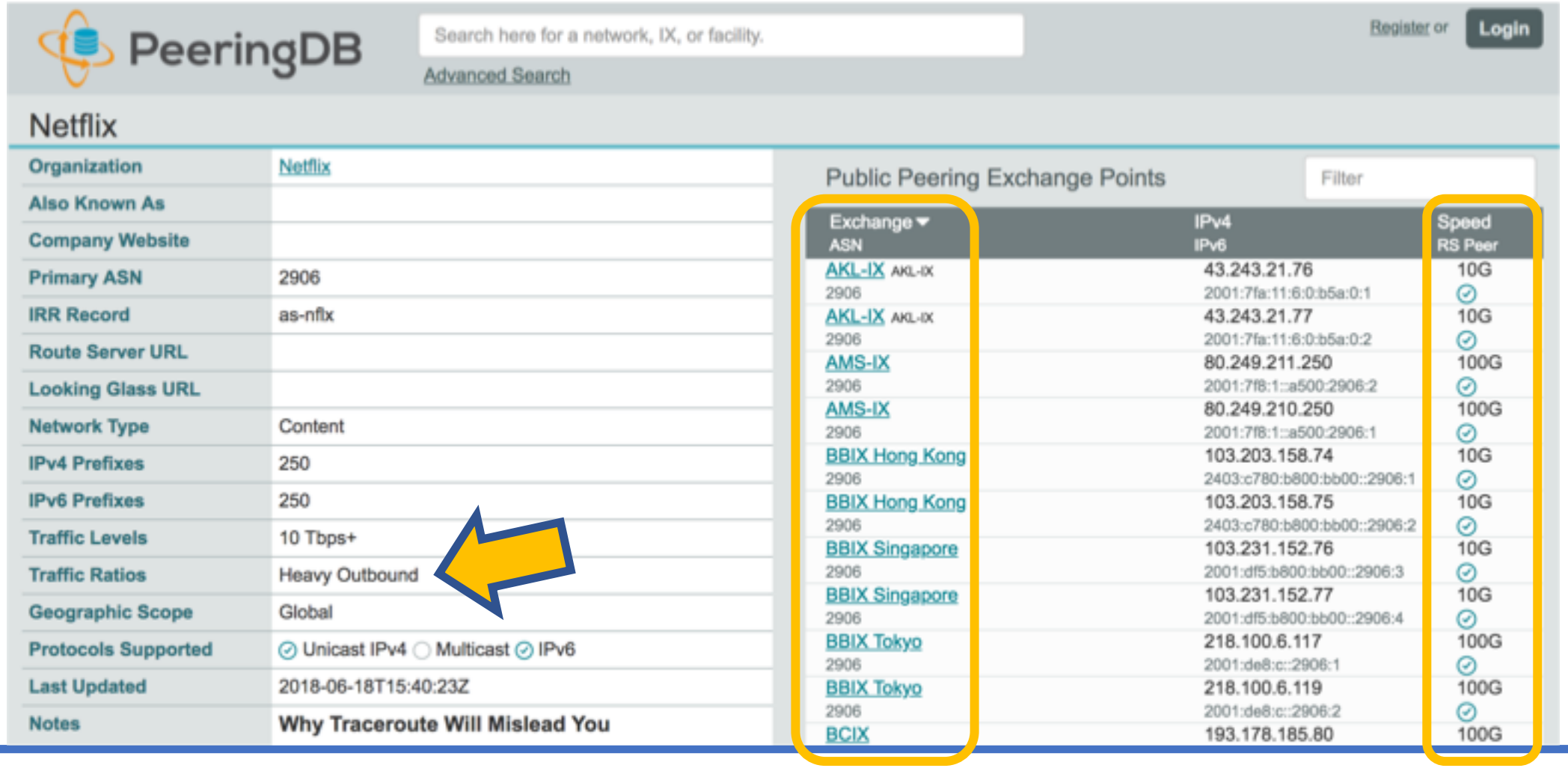
Queen Mary University of London

Hypergiants?

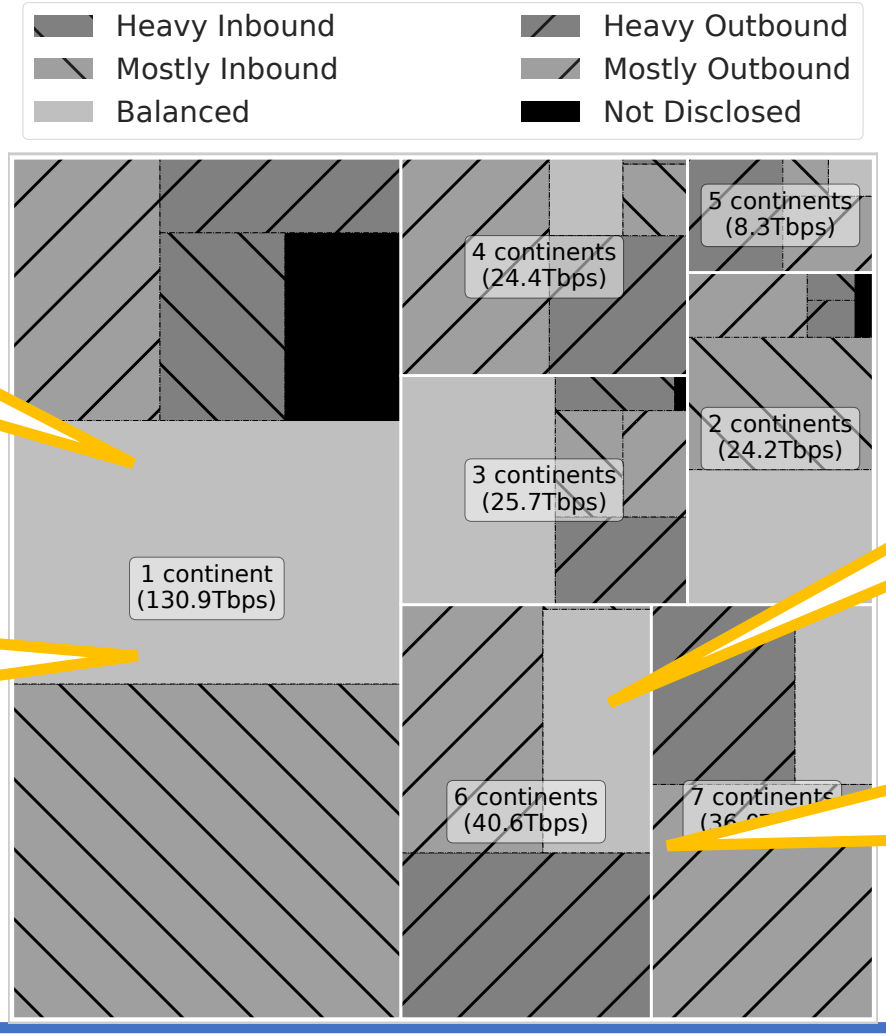
What would we need?

- Hypergiants are the ’biggest of the biggest’
  - Indication of **traffic volume**
- Hypergiants are global
  - Indication of **geographic reach**
- Hypergiants seem to be heavy on content
  - Indication of **traffic balance**

PeeringDB - Webinterface



Data distribution



Organisations at one continent: 45% of port capacity, yet 95% of all organisations

Organisations at one continent: mostly inbound or balanced traffic profile

Organisations at four continents or more : 38% of port capacity, yet only 1% of all organisations

Organisations at four continents or more : mostly outbound oriented

The problem

- Given the three features, classify whether an organisation is a hypergiant
  - Port capacity
  - Geographic reach
  - Traffic profile

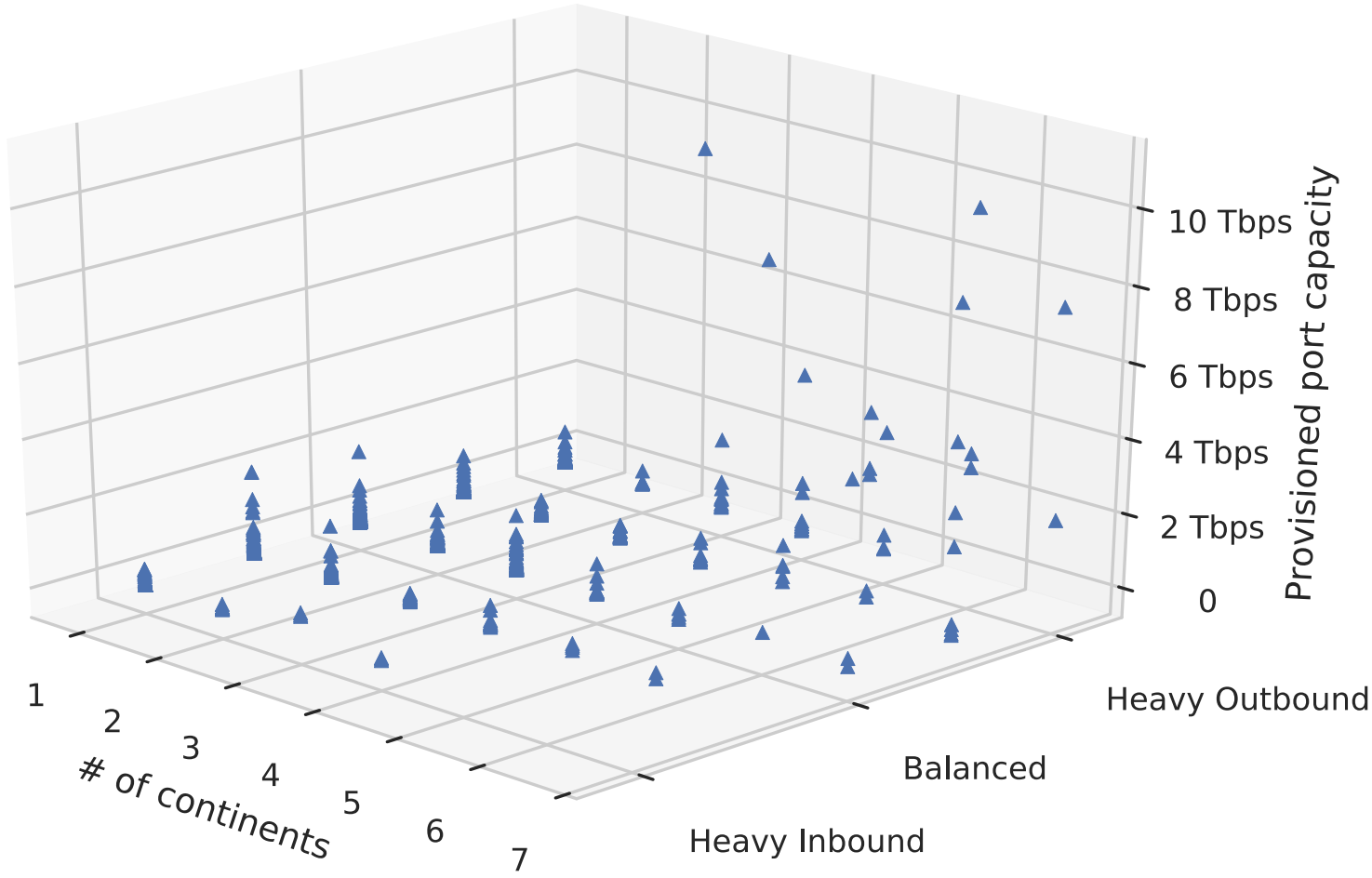
- Sounds like something that could be solved with ML techniques

- No labels, so no supervised learning

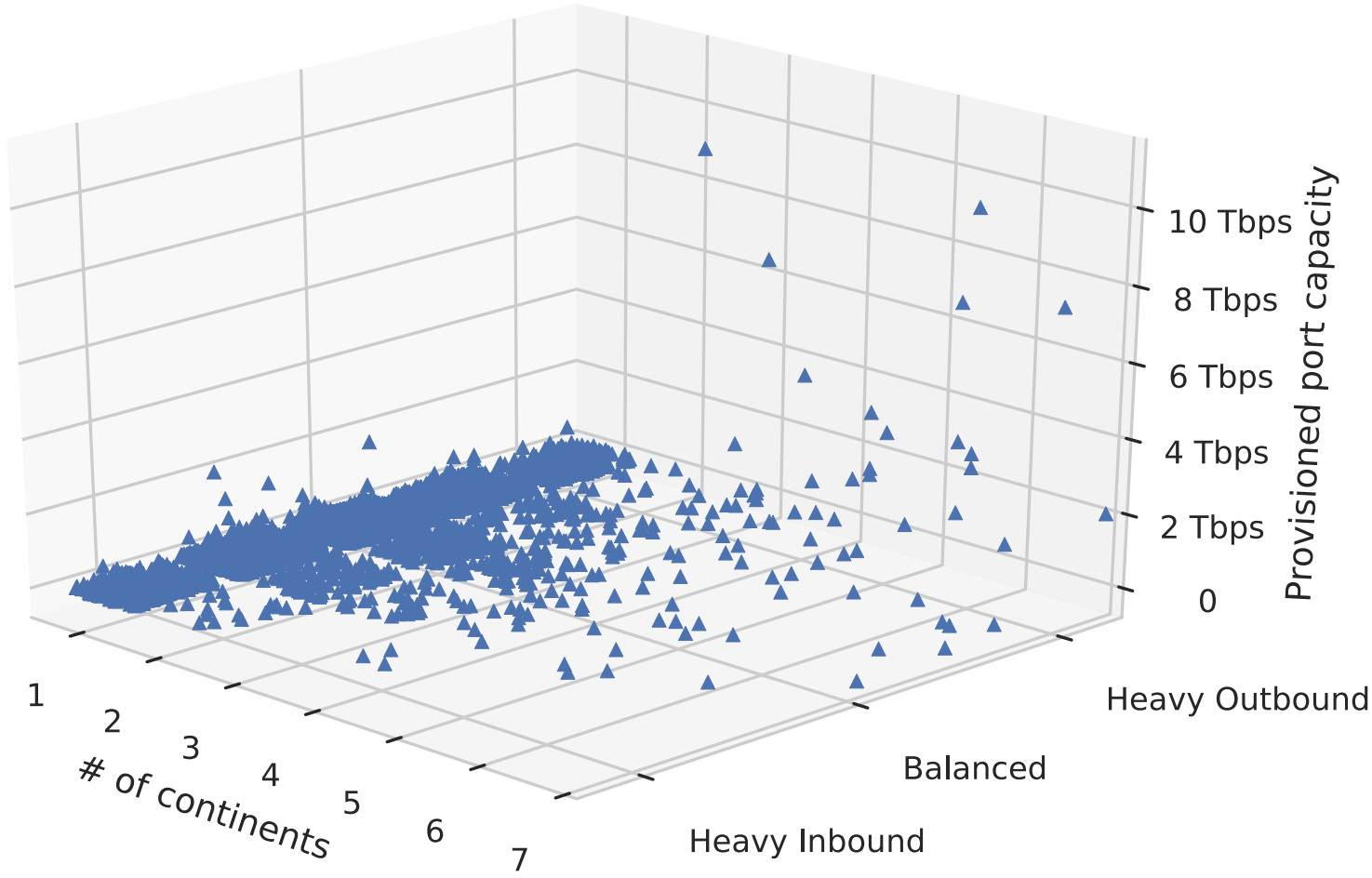
Intuition

- Hypergiants are ‘the biggest of the biggest’
- They must be different from the crowd somehow
  - On some metric the very least
- Use unsupervised learning
- Use k-means (with k=2) to actually get labelled data

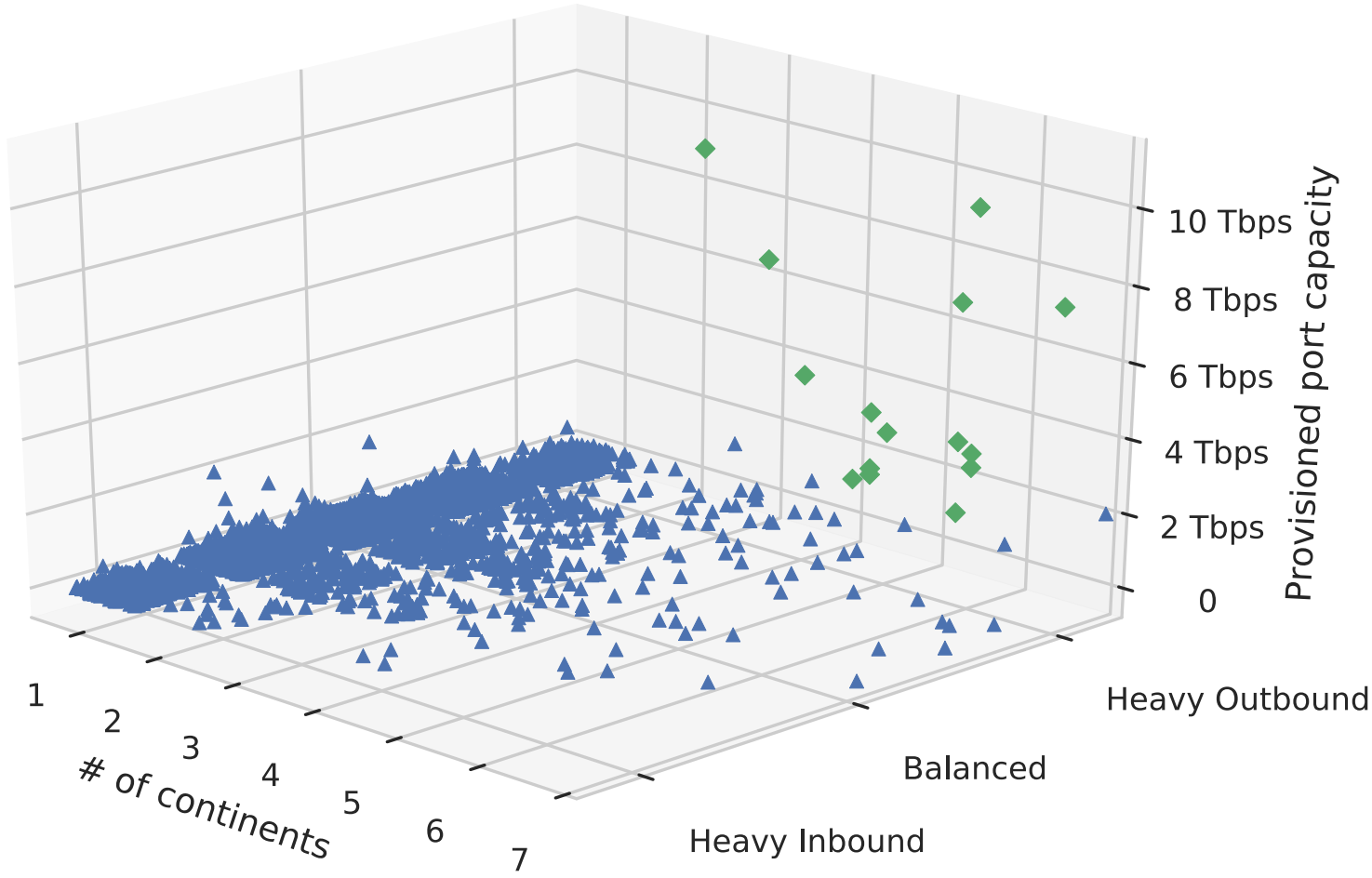
On the way…


10 Tbps
8 Tbps
6 Tbps
4 Tbps
2 Tbps
0
Provisioned port capacity
1
2
3
4
5
6
7
# of continents
Heavy Outbound
Balanced
Heavy Inbound

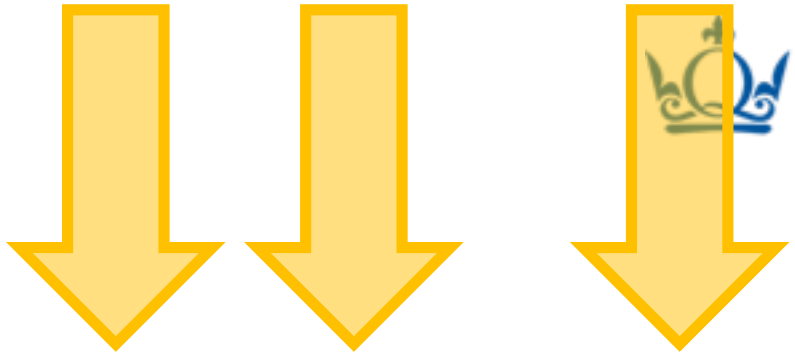
On the way…

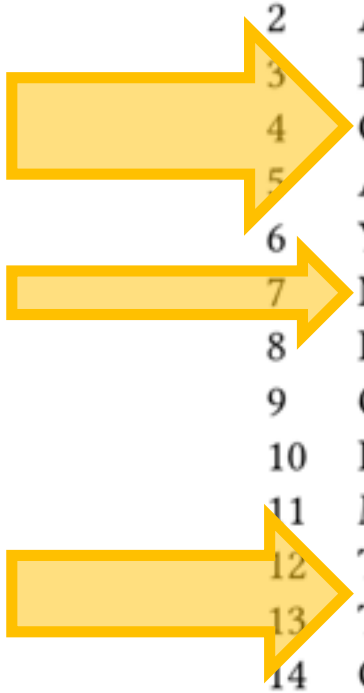

10 Tbps
8 Tbps
6 Tbps
4 Tbps
2 Tbps
0
Provisioned port capacity
1
2
3
4
5
6
7
# of continents
Heavy Inbound
Balanced
Heavy Outbound

Hypergiants!



Hypergiants!





	Organisation name	ASN	Continents	Port. Cap.	Traffic Profile
1	Apple Inc	714	4	10.960 Tbps	Mostly Outbound
2	Amazon.com	16509	6	9.991 Tbps	Balanced
3	Facebook	32934	6	9.840 Tbps	Heavy Outbound
4	Google Inc.	15169	7	8.741 Tbps	Mostly Outbound
5	Akamai Technologies	20940	7	7.854 Tbps	Heavy Outbound
6	Yahoo!	10310	6	5.310 Tbps	Mostly Outbound
7	Netflix	2906	7	5.170 Tbps	Mostly Outbound
8	Hurricane Electric	6939	7	5.037 Tbps	Balanced
9	OVH	16276	4	4.270 Tbps	Heavy Outbound
10	Limelight Networks Global	22822	6	3.840 Tbps	Mostly Outbound
11	Microsoft	8075	6	3.680 Tbps	Mostly Outbound
12	Twitter, Inc.	13414	6	3.401 Tbps	Heavy Outbound
13	Twitch	46489	5	3.340 Tbps	Heavy Outbound
14	Cloudflare	13335	7	3.320 Tbps	Mostly Outbound
15	Verizon Digital Media Services	15133	6	3.030 Tbps	Heavy Outbound

Summary

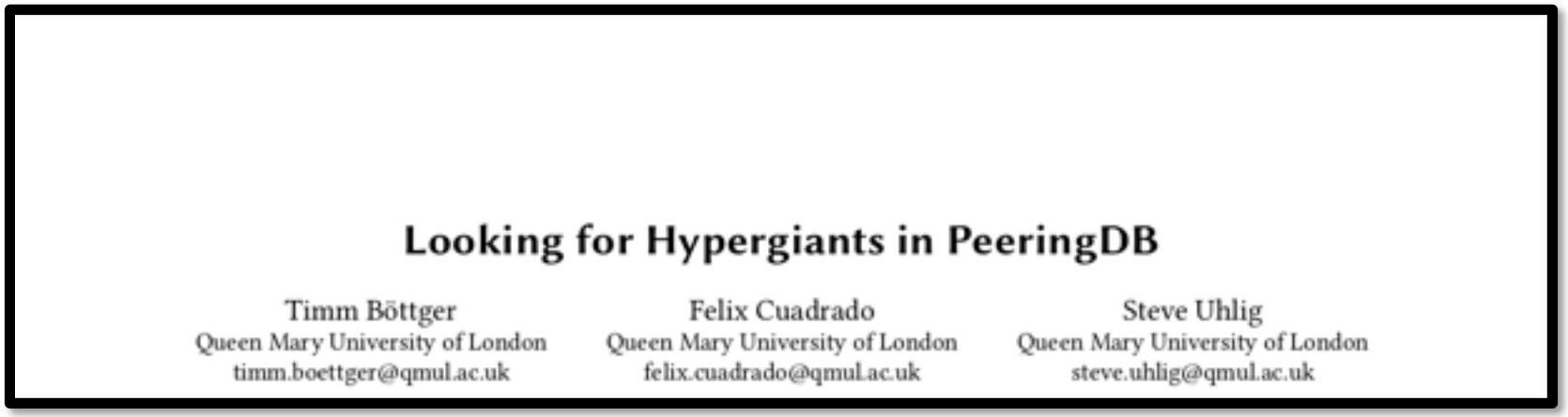
- Characterised organisations in PeeringDB
  - Provisioned port capacity
  - Geographic reach
  - Traffic profile

- Use unsupervised learning to differentiate hypergiants

- Shown that hypergiants actually are different enough to be identified by natural split in the data

Complete paper in CCR

- Complete paper with all results to appear in the July issue of ACM SIGCOMM CCR.



Looking for Hypergiants in PeeringDB

Timm Böttger
Queen Mary University of London
timm.boettger@qmul.ac.uk

Felix Cuadrado
Queen Mary University of London
felix.cuadrado@qmul.ac.uk

Steve Uhlig
Queen Mary University of London
steve.uhlig@qmul.ac.uk

Queen Mary University of London

Thank you very much for listening!

Questions?