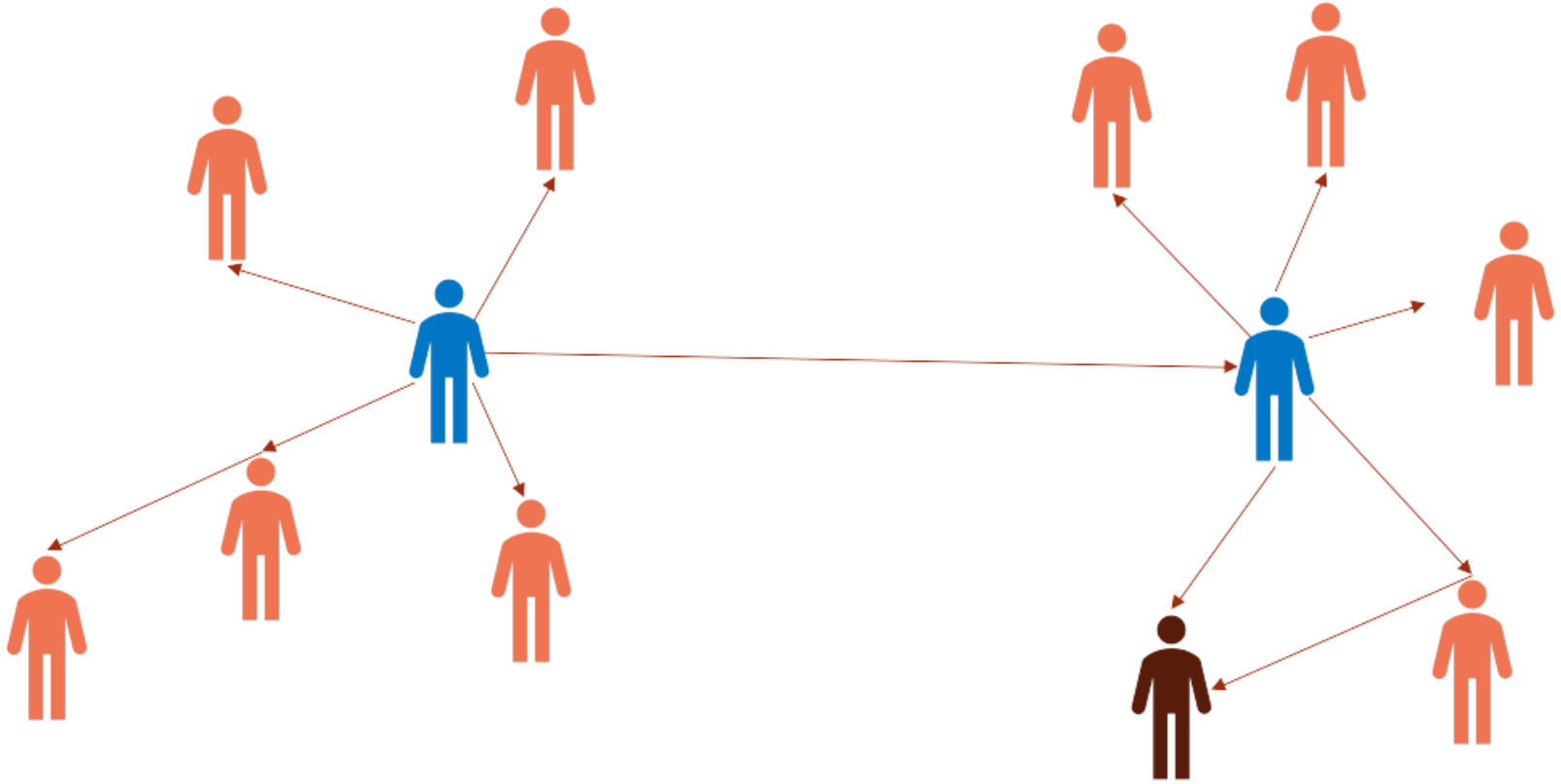
INFERRING THE GLOBAL COVID-19 TRANSMISSION NETWORK USING TNET

Saurav Dhar, Ion Mandoiu, Mukul S. Bansal

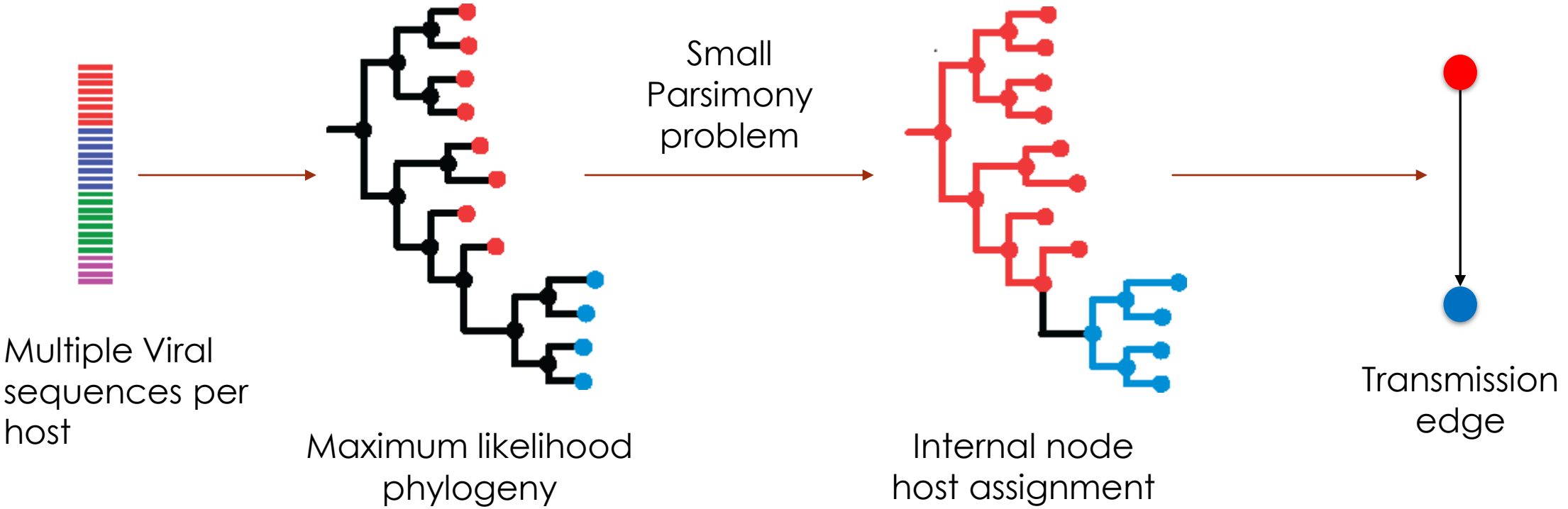
TRANSMISSION NETWORK

Who infected whom?

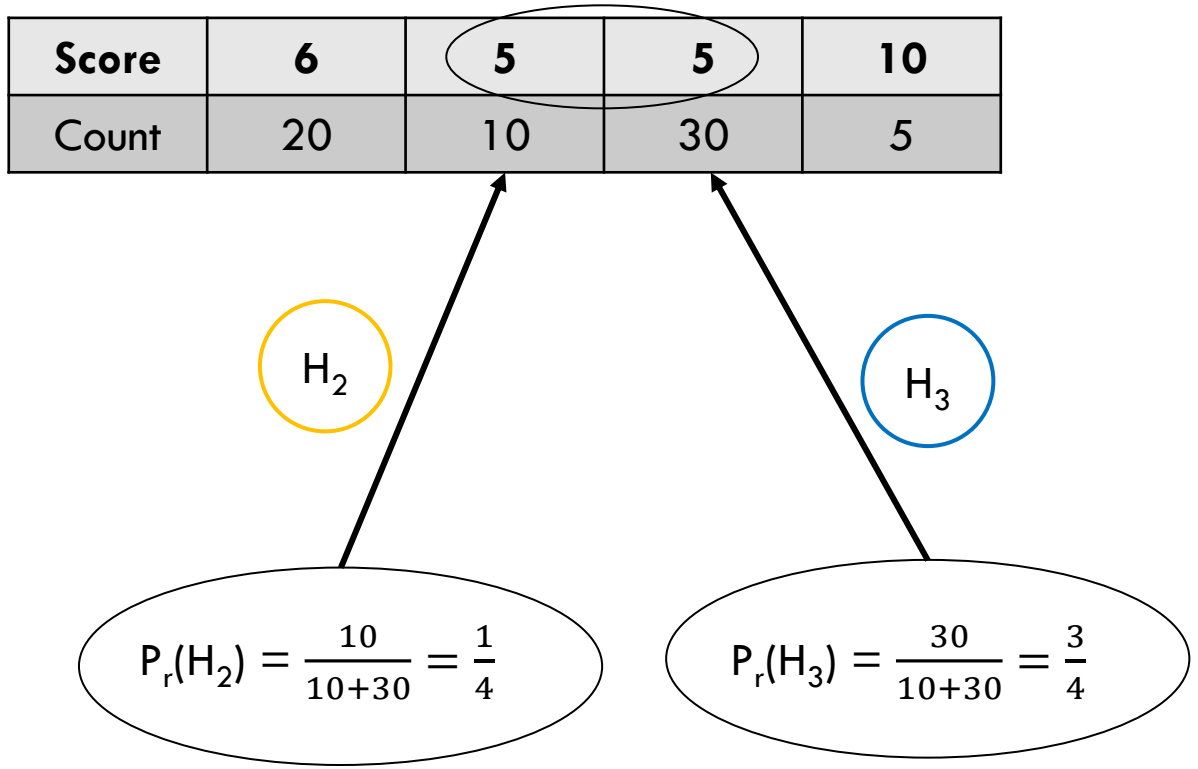
Multiple viral sequences sampled from the hosts.



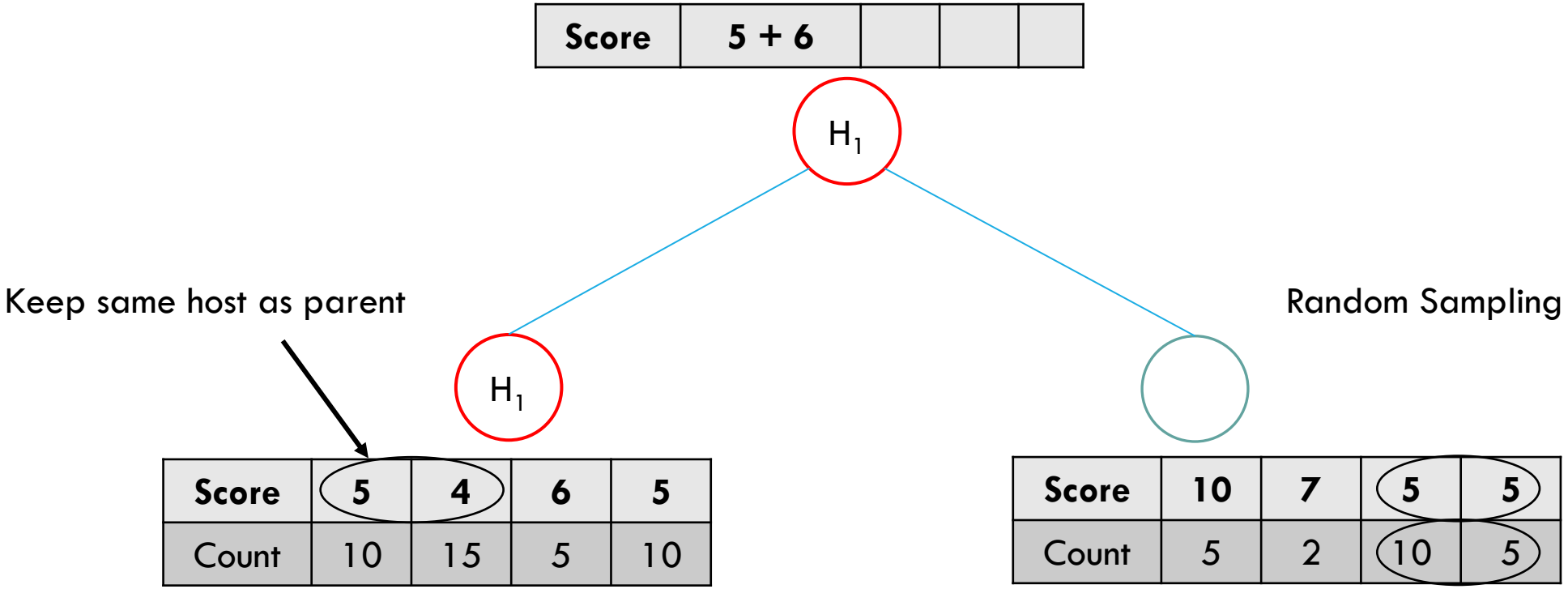
TNET WORKFLOW



DEALING WITH MULTIPLE OPTIMAL HOST ASSIGNMENT IDEA 1: RANDOM SAMPLING

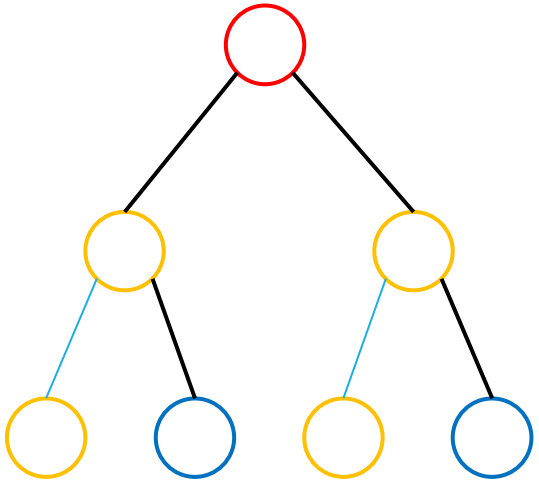


DEALING WITH MULTIPLE OPTIMAL HOST ASSIGNMENT IDEA 2: BIASED SAMPLING

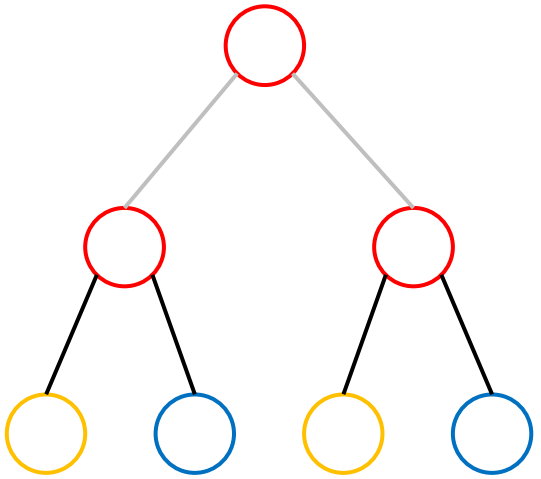


TNET-RANDOM VS. TNET-BIASED

Parsimony score = 4

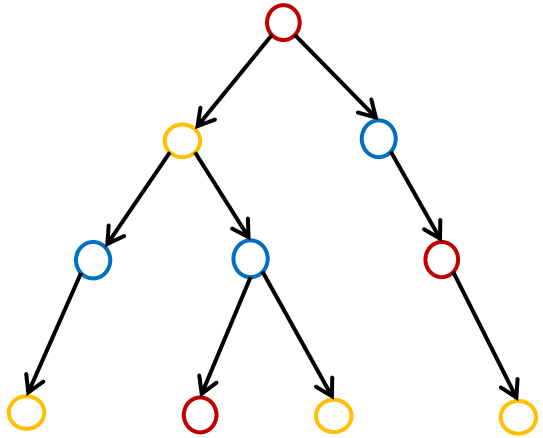


Parsimony score = 4

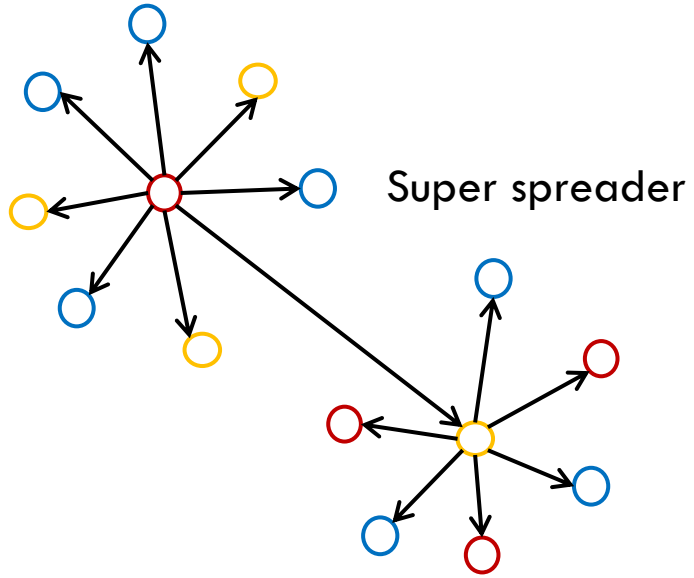


Changes are pushed further down the tree keeping the same parsimony score

TNET-RANDOM VS. TNET-BIASED: RELATIVE ACCURACY DEPENDS ON NETWORK



Infector ratio ~ 2



Infector ratio ~ 7

Based on simulated data, we see that TNet-Biased works better for networks with higher infector ratio.

TNET ON COVID-19 GLOBAL COUNTRY LEVEL ANALYSIS

Dataset

- 59 countries
- 2123 total complete sequences
- 10 to 100 sequences per country
- Up to June, 2020

Measure

- Country of exposure
- Top 5 spreaders and receivers in each month
- Top 5 spreaders and receivers of USA

Compared against NextStrain

- True country of exposure known

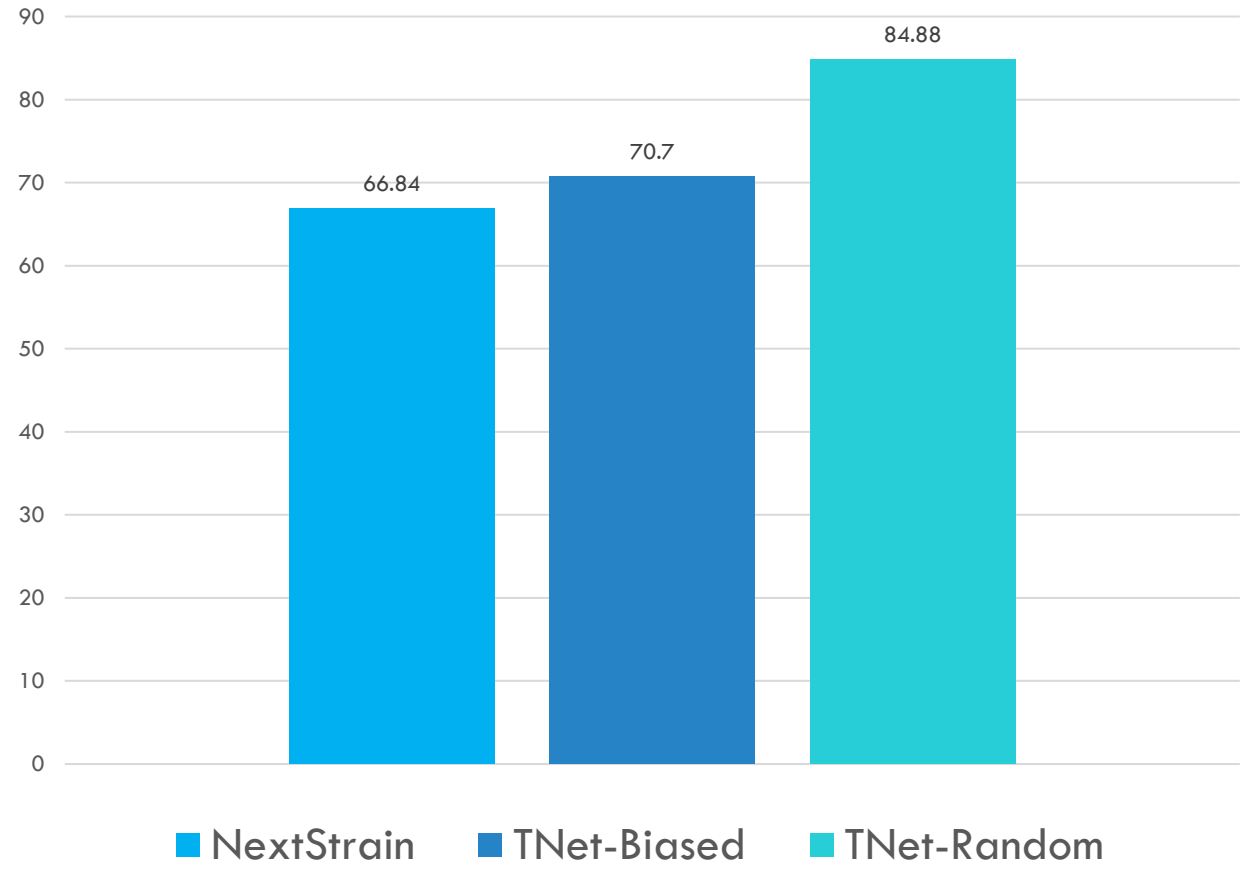
Handling topological uncertainty

- 10 bootstrap trees

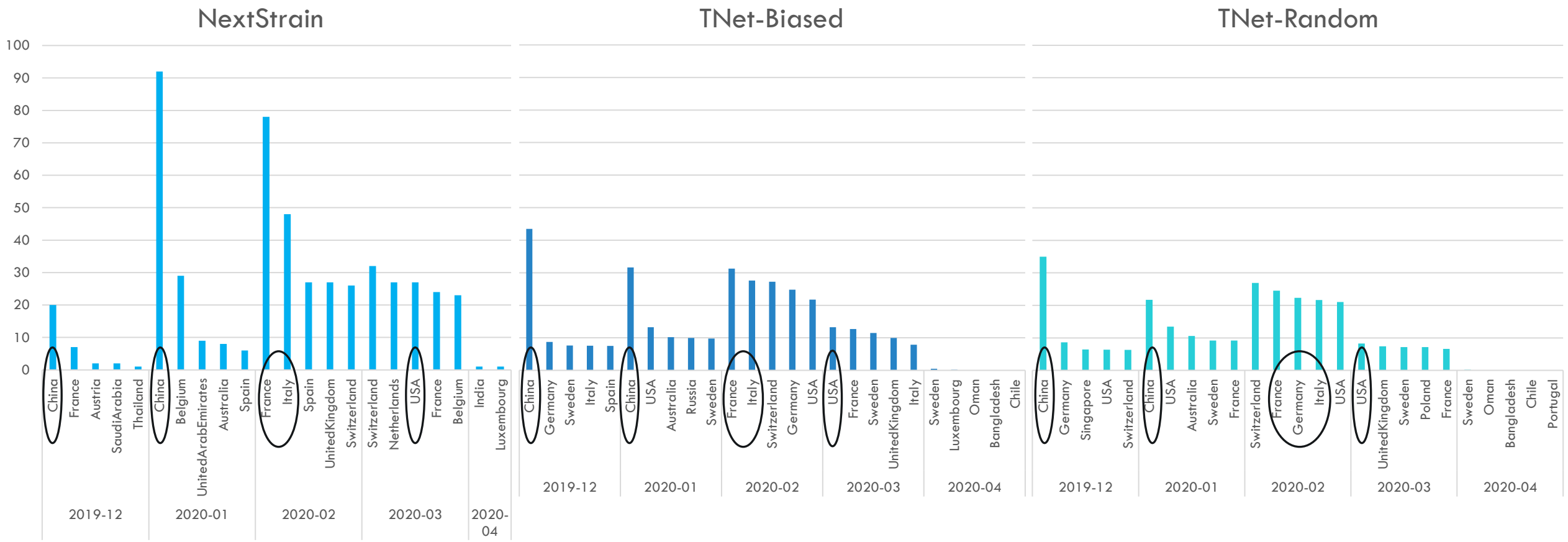
Handling sampling uncertainty

- Used 100 samples per bootstrap tree

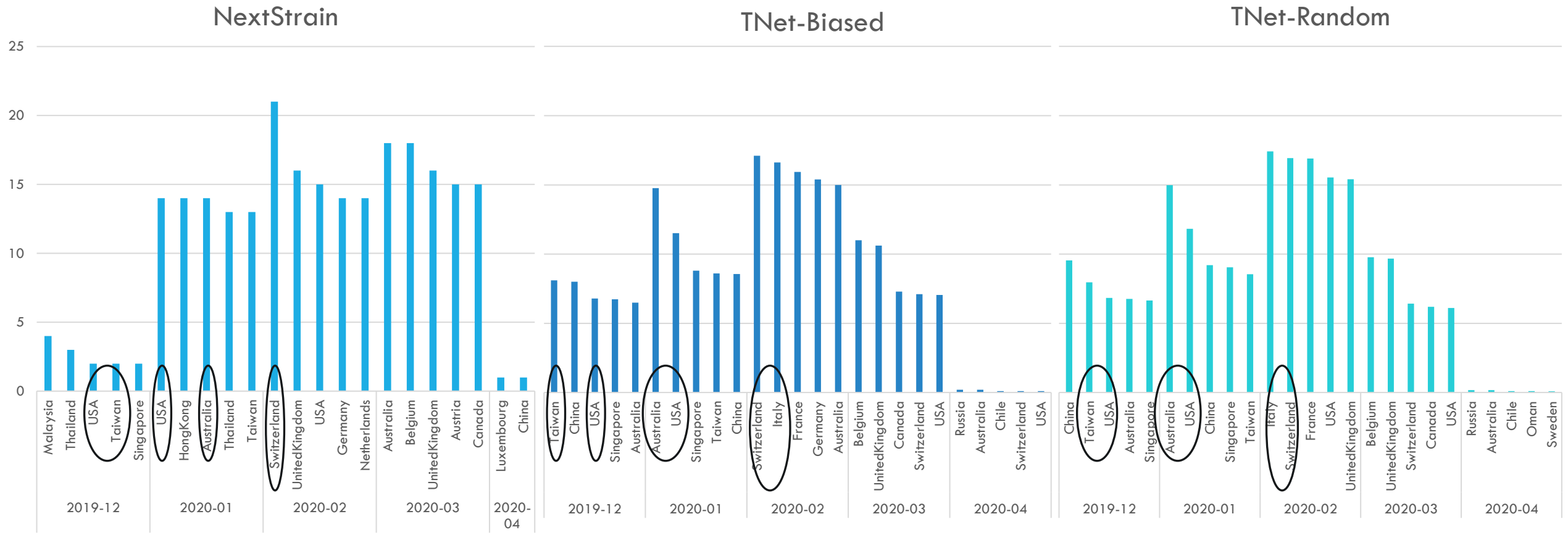
ACCURACY: KNOWN COUNTRY OF EXPOSURE FOR EACH SEQUENCE



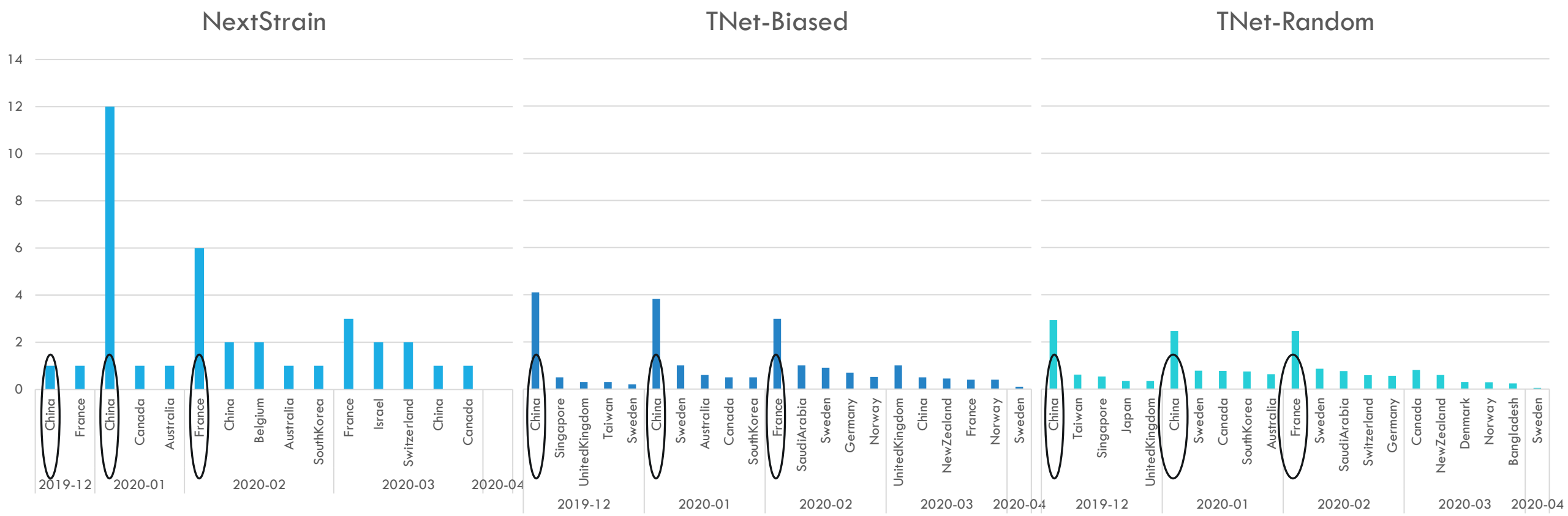
TOP 5 SPREADERS OF EACH MONTH (AVERAGED)



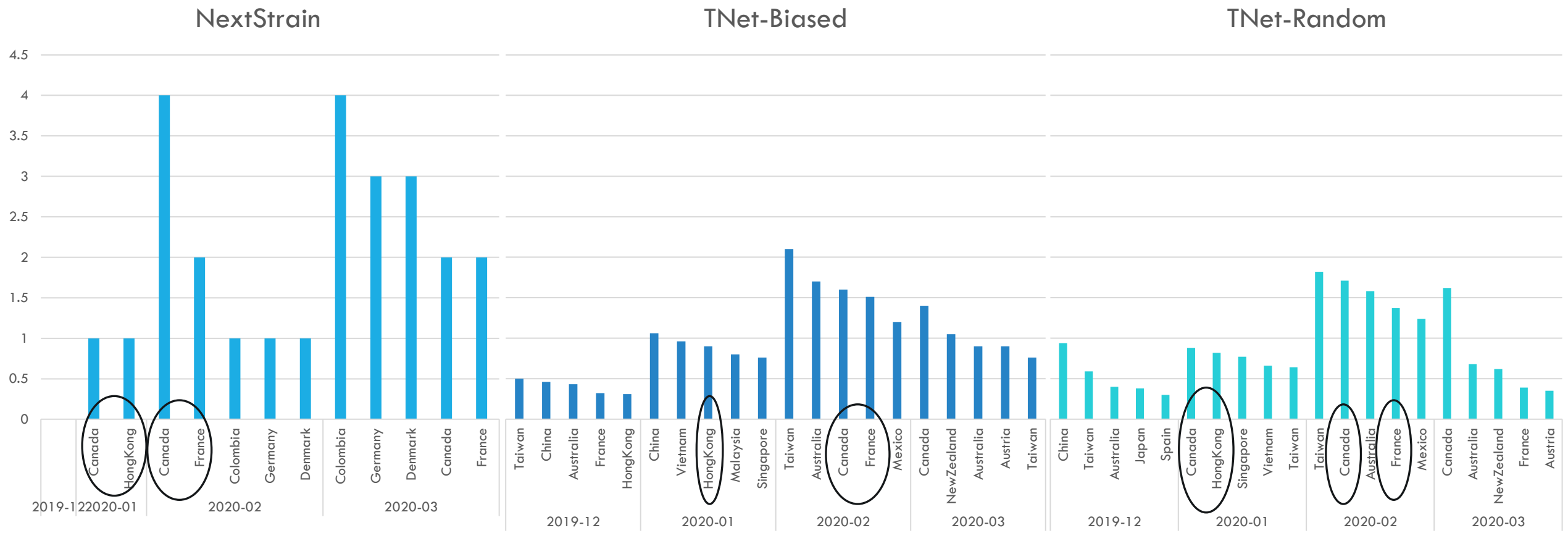
TOP 5 RECEIVERS OF EACH MONTH (AVERAGED)



SPREAD TO USA: TOP 5 COUNTRIES



SPREAD FROM USA: TOP 5 COUNTIES



TNET ON COVID-19 STATE LEVEL ANALYSIS FOR USA

Dataset

- 30 states
- 1801 total complete sequences
- 10 to 100 sequences per state
- Up to July, 2020

Measure

- State of exposure
- Top 5 spreaders and receivers in each month
- Top 5 spreaders and receivers of New York

Compared against NextStrain

- True state of exposure known

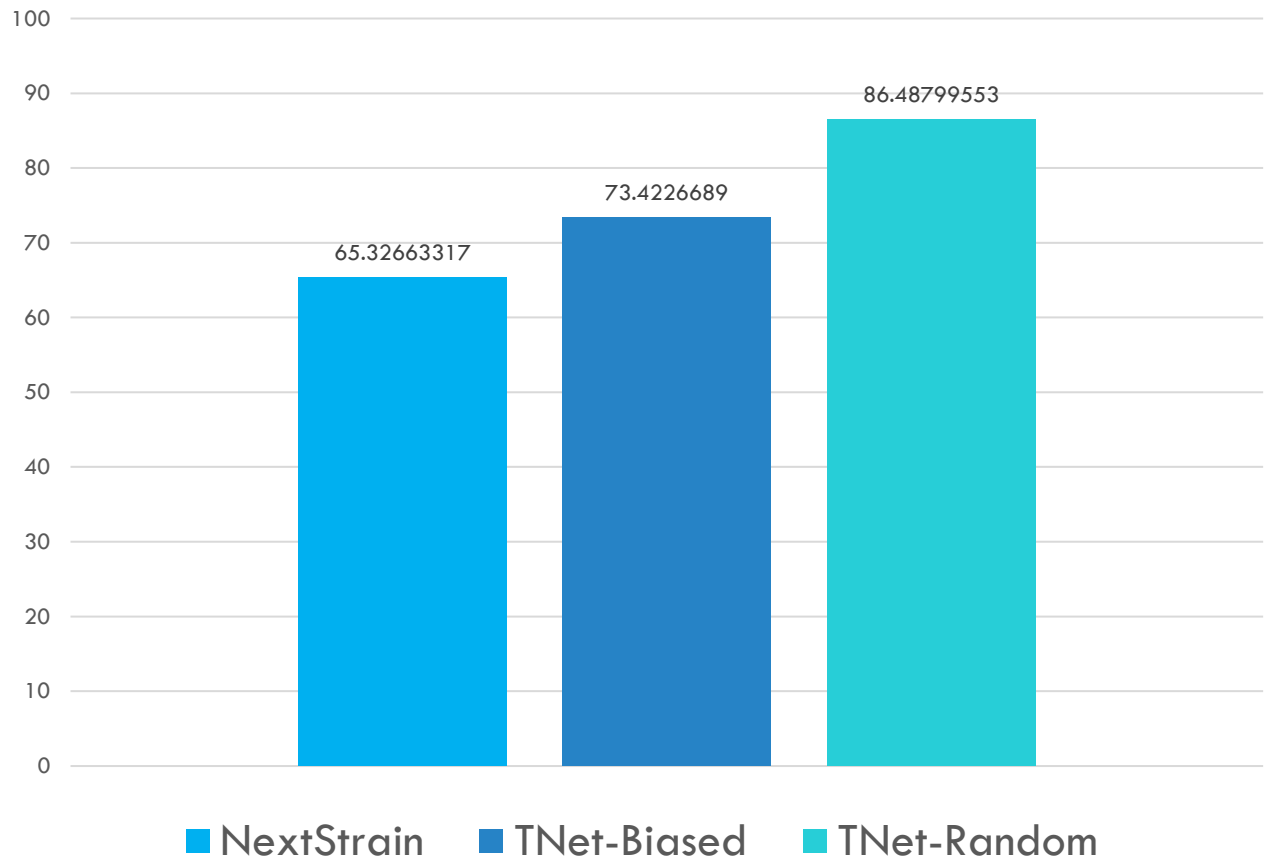
Handling topological uncertainty

- 10 bootstrap trees

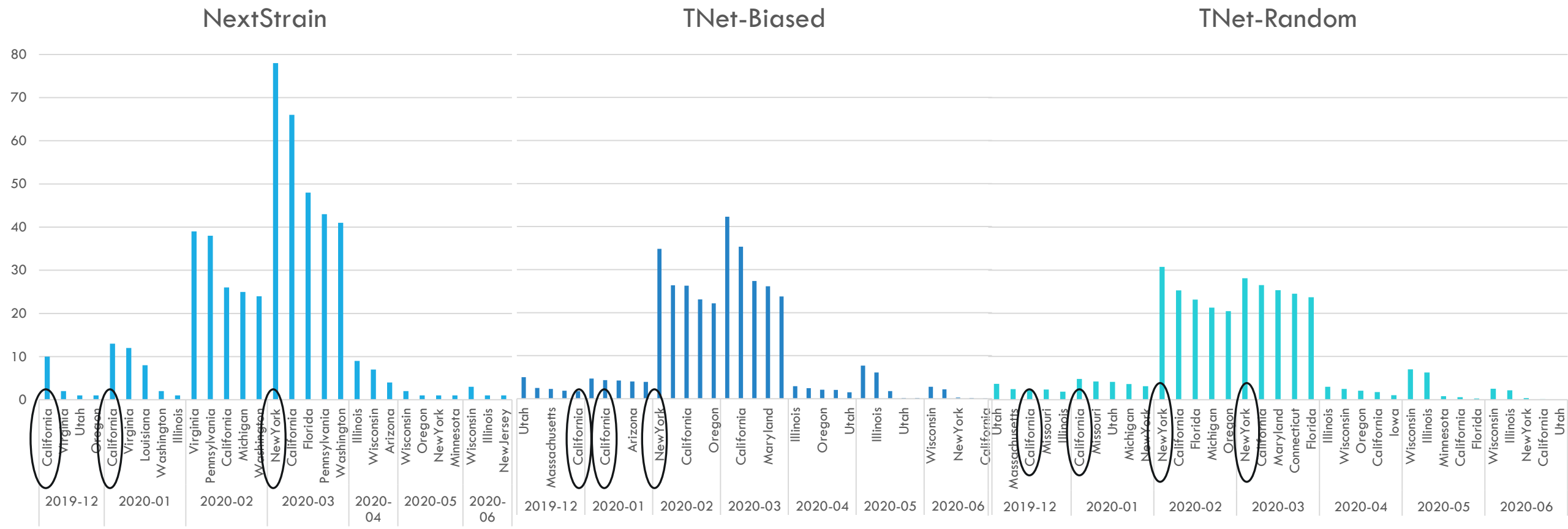
Handling sampling uncertainty

- Used 100 samples per bootstrap tree

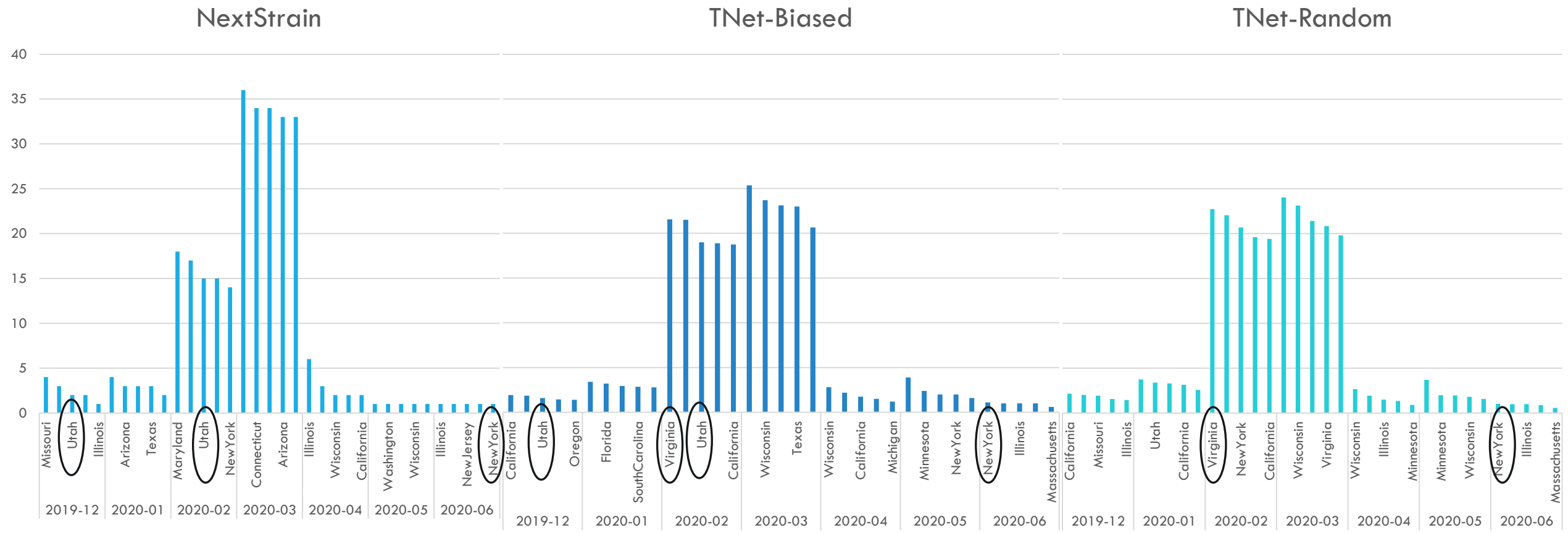
ACCURACY : KNOWN STATE OF EXPOSURE FOR EACH SEQUENCE



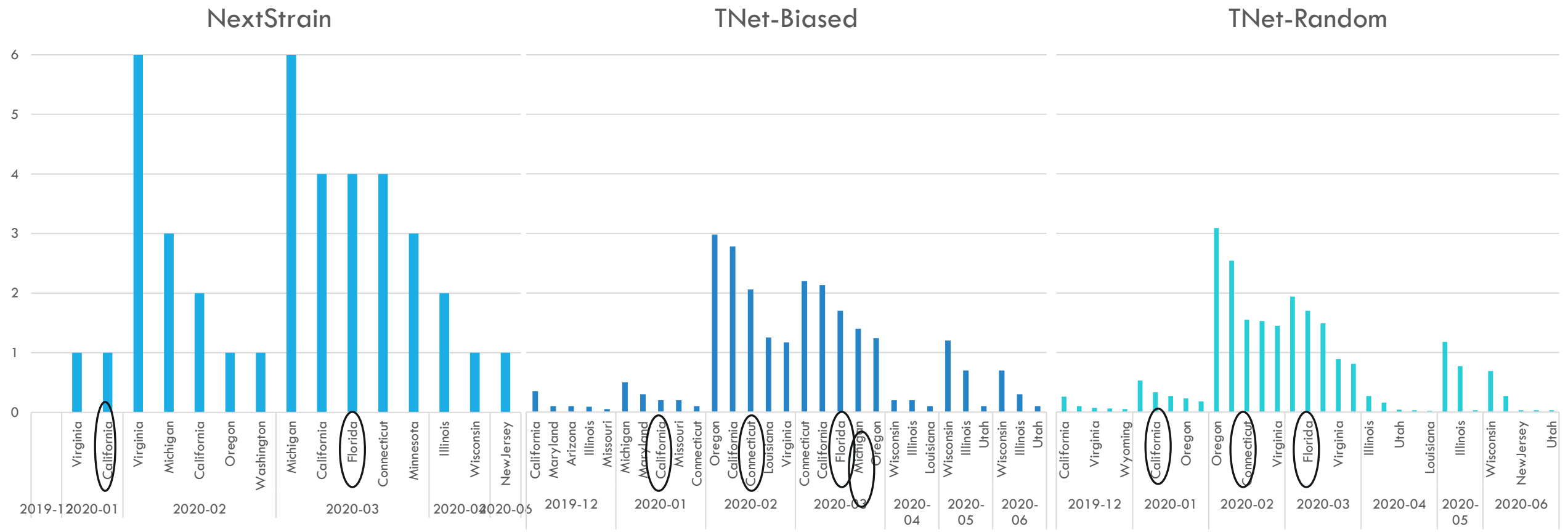
TOP 5 SPREADERS OVER TIME (AVERAGE)



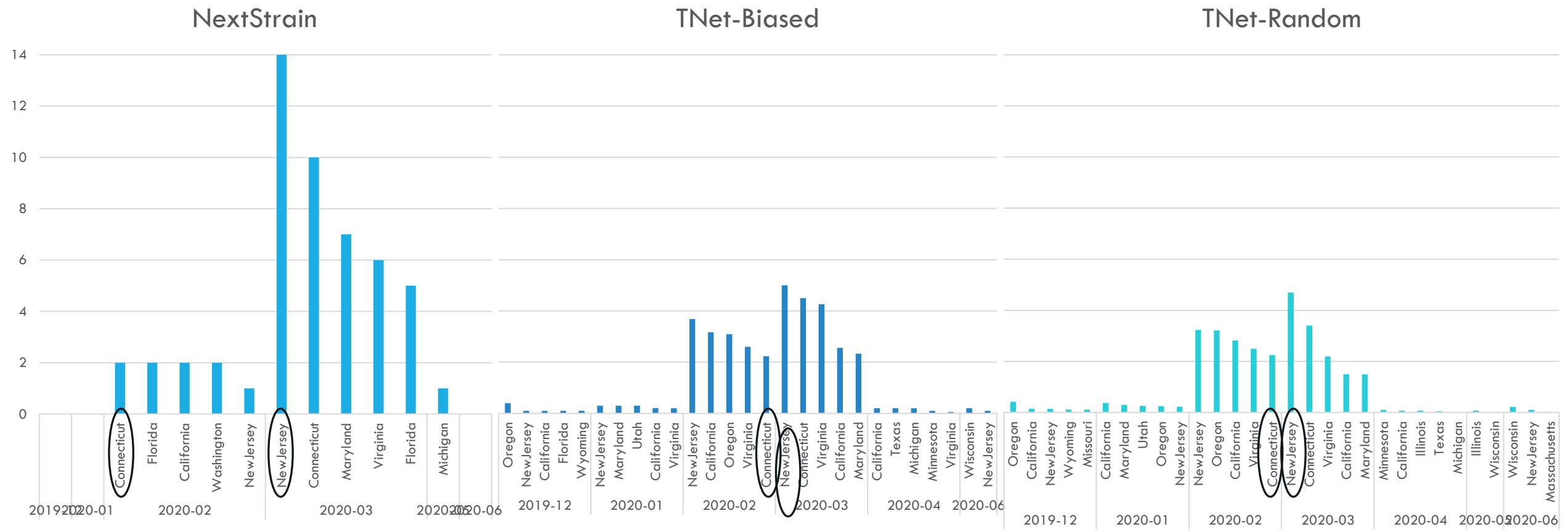
TOP 5 RECEIVERS OVER TIME (AVERAGE)



TOP 5 STATES NEW YORK RECEIVED FROM (AVERAGE)



TOP 5 STATES NEW YORK SPREAD TO (AVERAGE)

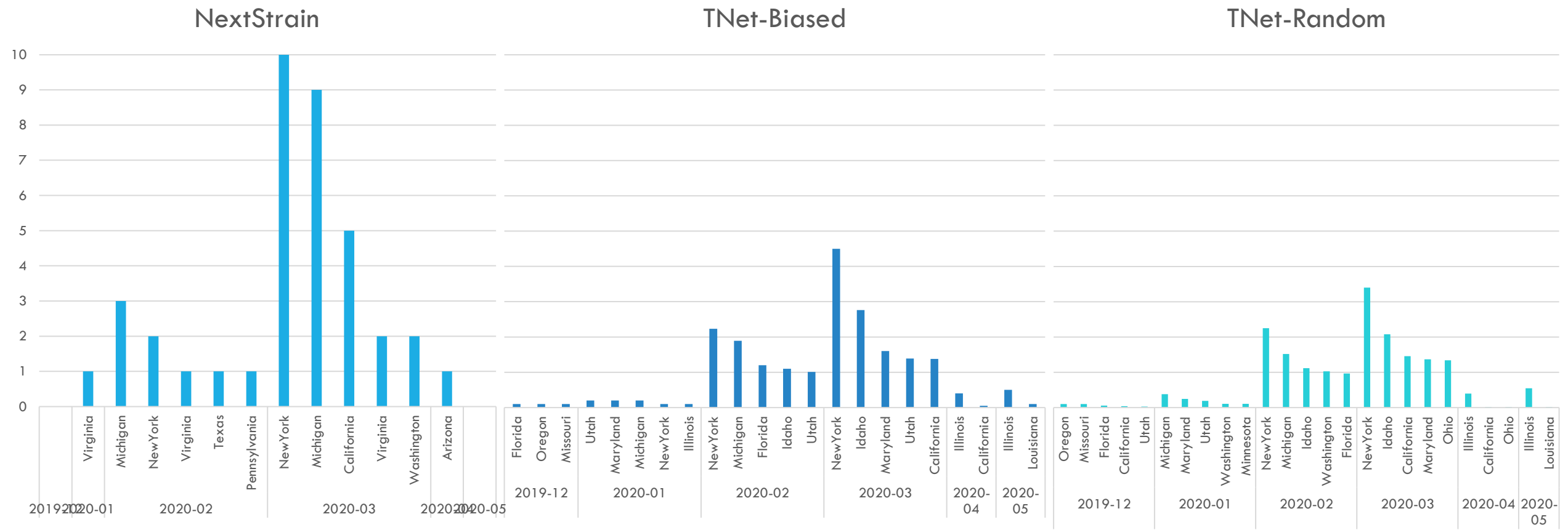


CONCLUSION

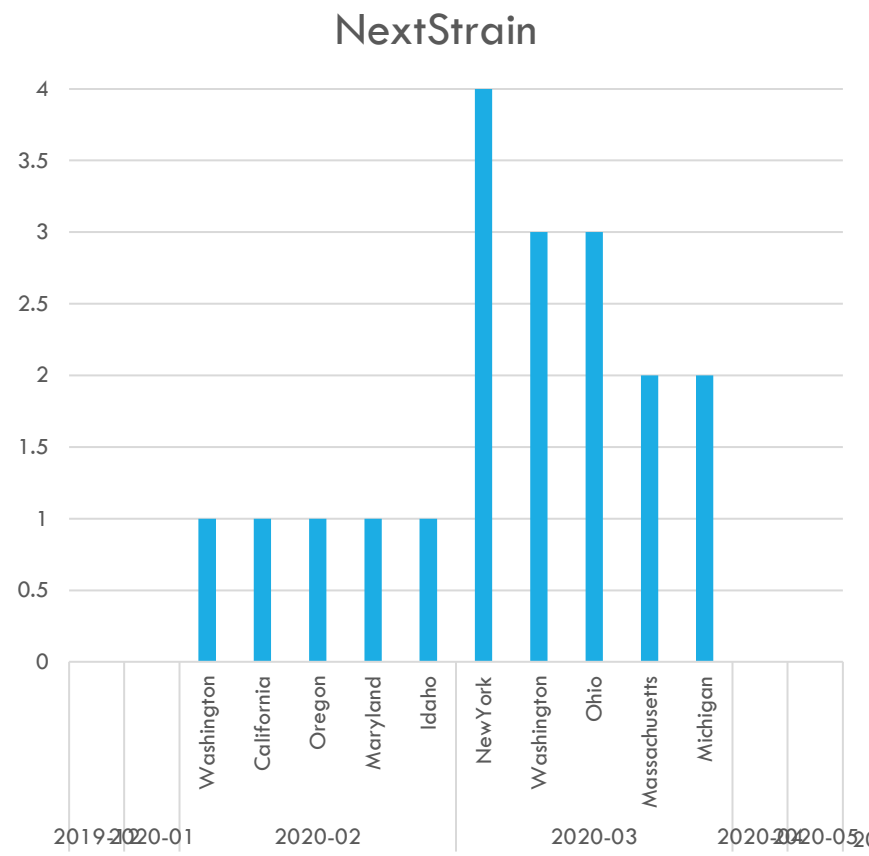
- Simple and scalable model like TNet-Random appears to work very well.
- Using countries and states as hosts may be a viable approach for such geographic spread analyses.

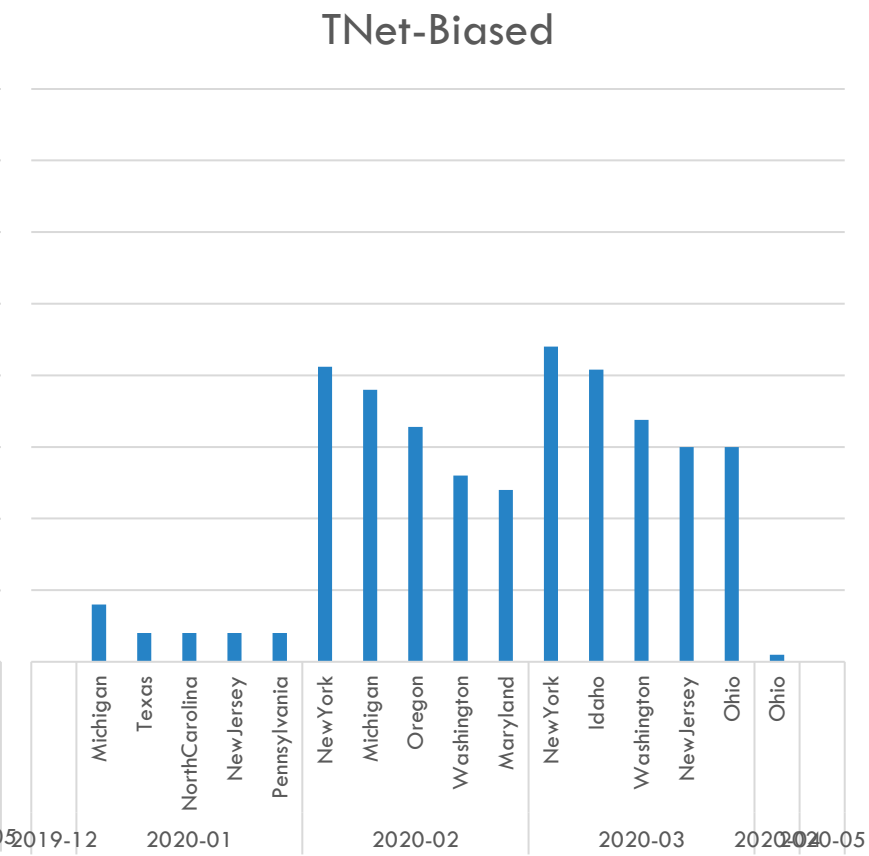

THANK YOU

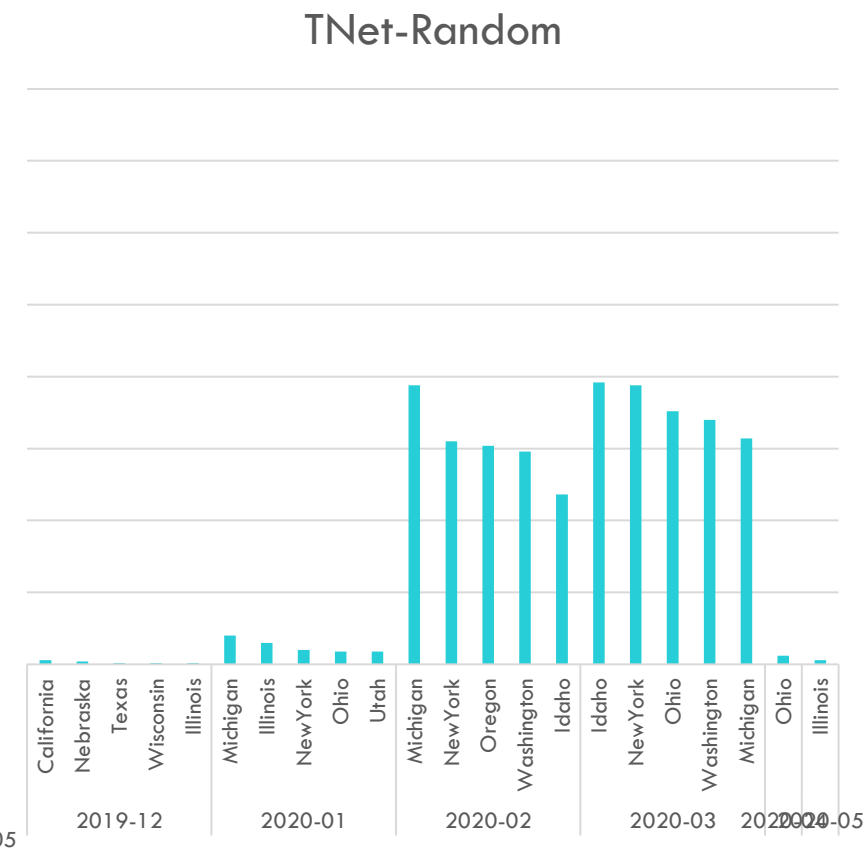
TOP 5 STATES CONNECTICUT RECEIVED FROM (AVERAGE)



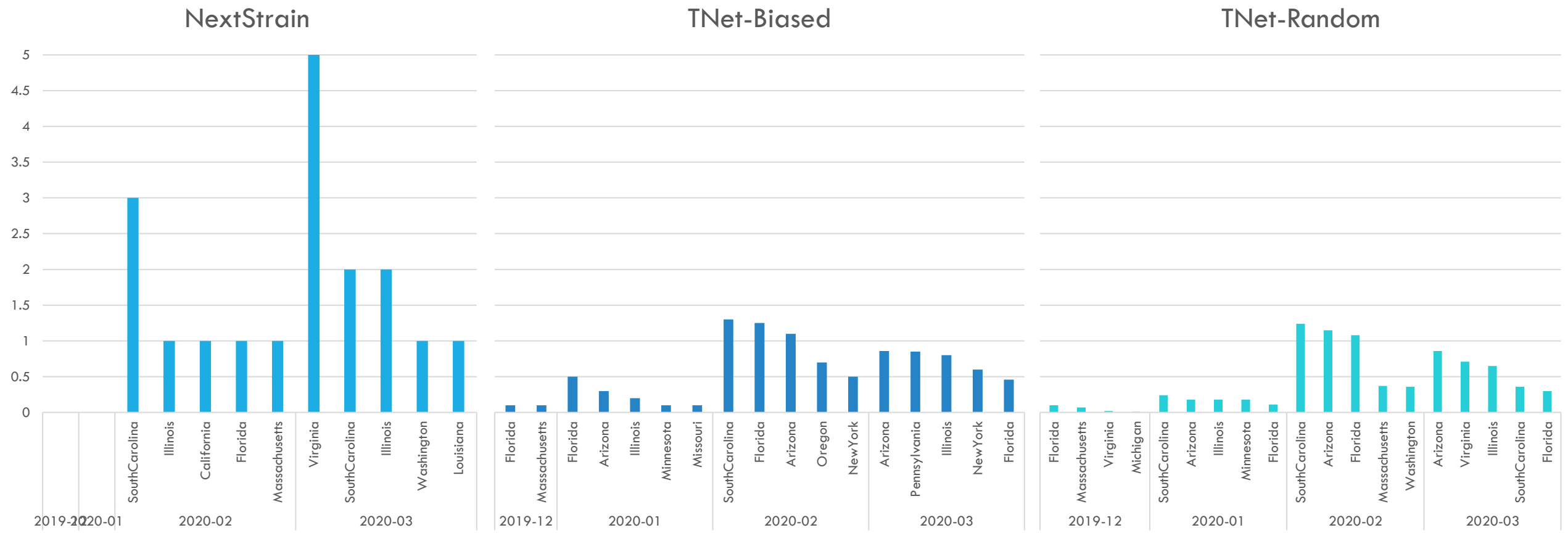
TOP 5 STATES CONNECTICUT SPREAD TO (AVERAGE)







TOP 5 STATES GEORGIA RECEIVED FROM (AVERAGE)



TOP 5 STATES GEORGIA SPREAD TO (AVERAGE)