Explaining Wide Area Data Transfer Performance

Zhengchun Liu, Prasanna Balaprakash, Rajkumar Kettimuthu and Ian Foster
(Argonne National Laboratory, Lemont, IL, USA)

Argonne NATIONAL LABORATORY

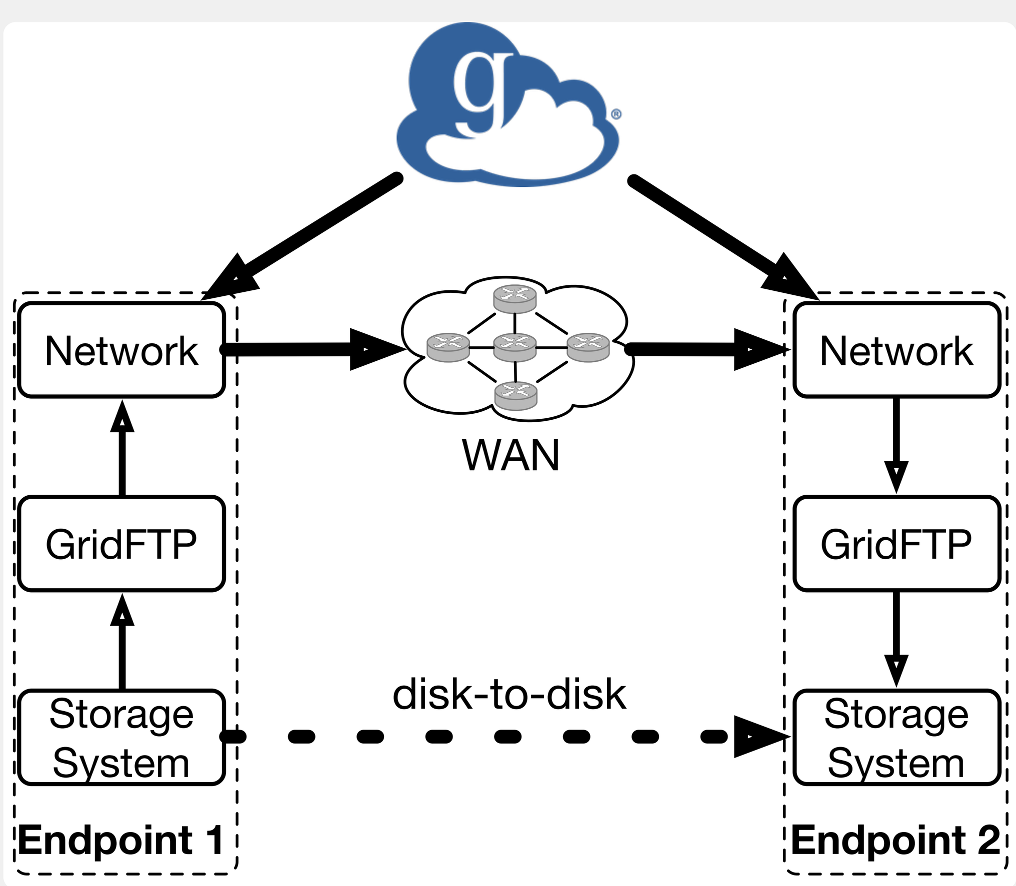
26th HPDC

Summary

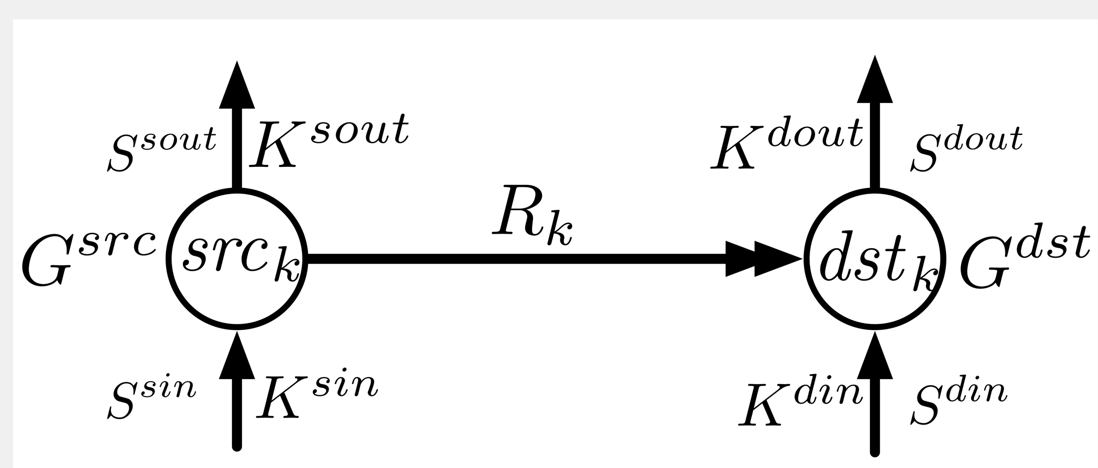
- Disk-to-disk wide-area file transfers involve many subsystems and tunable application parameters that pose significant challenges for bottleneck detection, system optimization, and performance prediction.
- Starting with log data for millions of Globus transfers involving billions of files and hundreds of petabytes, we engineer features for endpoint CPU load, network interface card load, and transfer characteristics; and we use these features in both linear and nonlinear models of transfer performance.
- By using our features, for a representative set of 30,653 transfers over 30 heavily used source-destination pairs ("edges"), totaling 2,053 TB in 46.6 million files, we obtain median absolute percentage prediction errors (MdAPE) of 7.0% and 4.6% when using distinct linear and nonlinear models per edge, respectively;
- We found that the contention at endpoints can significantly reduce aggregate performance of even overprovisioned networks. This result suggests that aggregate performance can be improved by scheduling transfers and/or reducing concurrency and parallelism.
- The feature engineering work provides useful hints and insights for data science practitioners in wide area data transfer. It broadens understanding of factors that influence file transfer rate by clarifying relationships between achieved transfer rates, transfer characteristics, and competing load.

1. Quantifying resource contention

Characterize the resource contention of an end-to-end file transfer

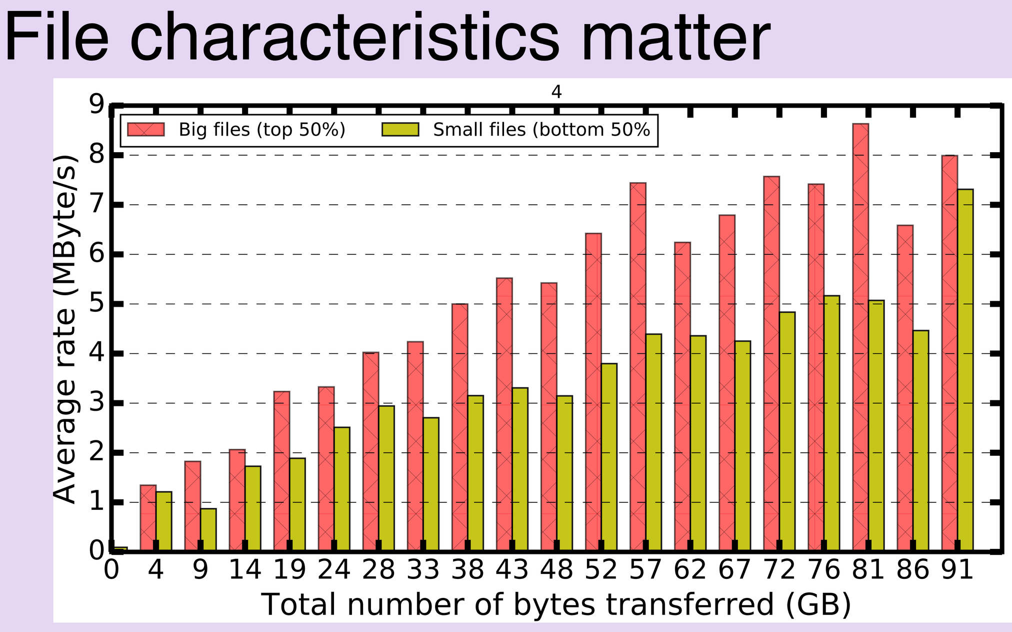


Structure of a Globus end-to-end file transfer from source (left) to destination (right), managed by cloud service.



Load experienced by a Globus transfer k from endpoint src^k to endpoint dst^k.

File characteristics matter



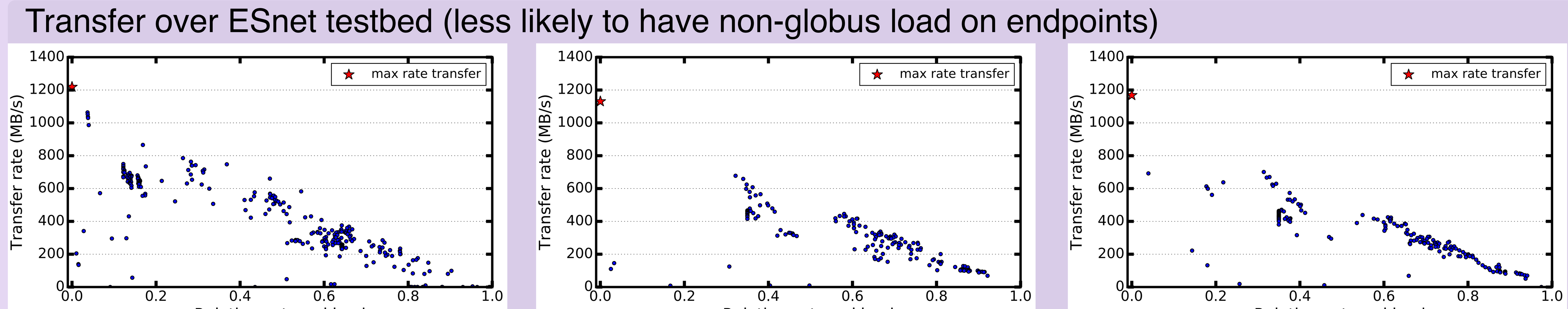
Features to explain a transfer

K^{sin}	Contending incoming transfer rate on src_k.
K^{sout}	Contending outgoing transfer rate on src_k.
K^{din}	Contending incoming transfer rate on dst_k.
K^{dout}	Contending outgoing transfer rate on dst_k.
C	Concurrency: Number of GridFTP processes.
P	Parallelism: Number of TCP channels per process.
S^{sin}	Number of incoming TCP streams on src_k.
S^{sout}	Number of outgoing TCP streams on src_k.
S^{din}	Number of incoming TCP streams on dst_k.
S^{dout}	Number of outgoing TCP streams on dst_k.
G^{src}	GridFTP instance count on src_k.
G^{dst}	GridFTP instance count on dst_k.
Nf	Number of files transferred.
Nd	Number of directories transferred.
Nb	Total number of bytes transferred.

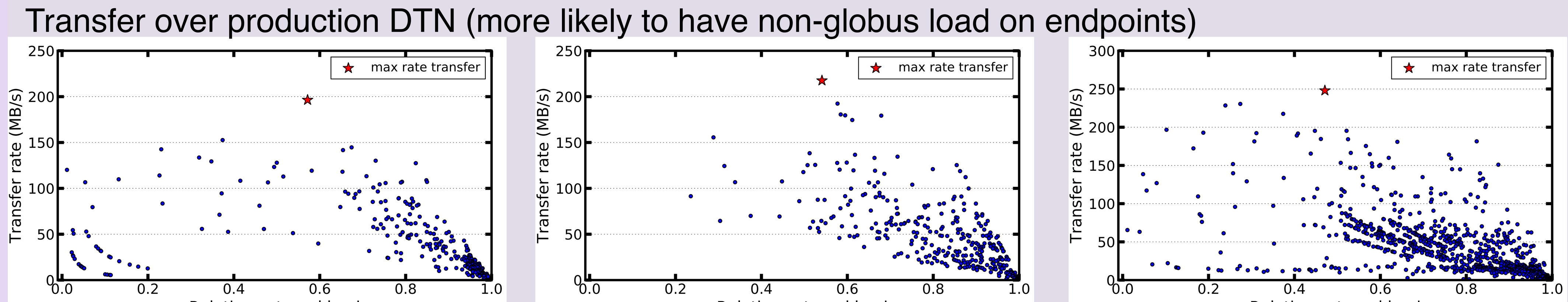
2. Factors that affect transfer performance

Contentions from non-globus programs on endpoint also matter

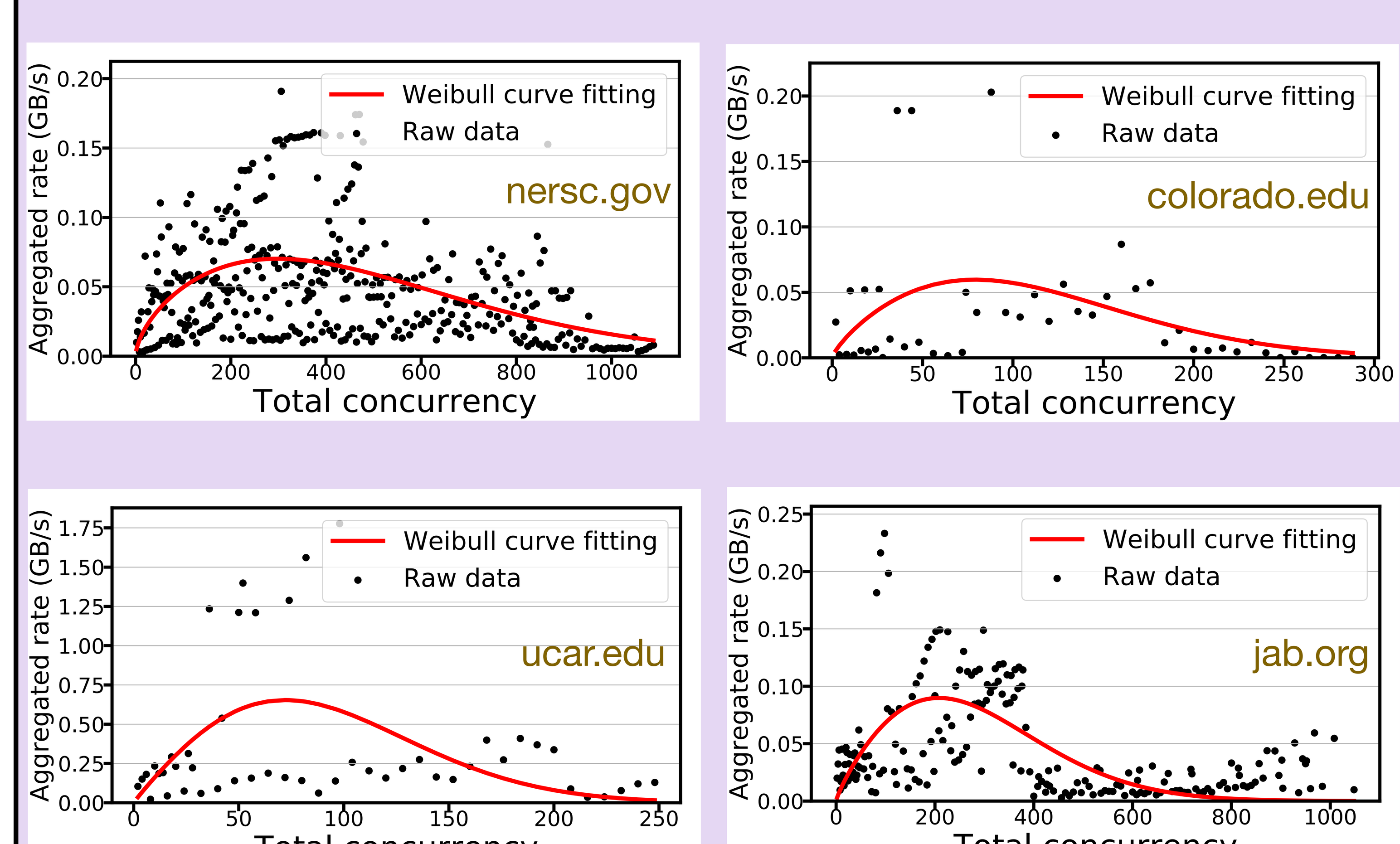
Transfer over ESnet testbed (less likely to have non-globus load on endpoints)



Transfer over production DTN (more likely to have non-globus load on endpoints)

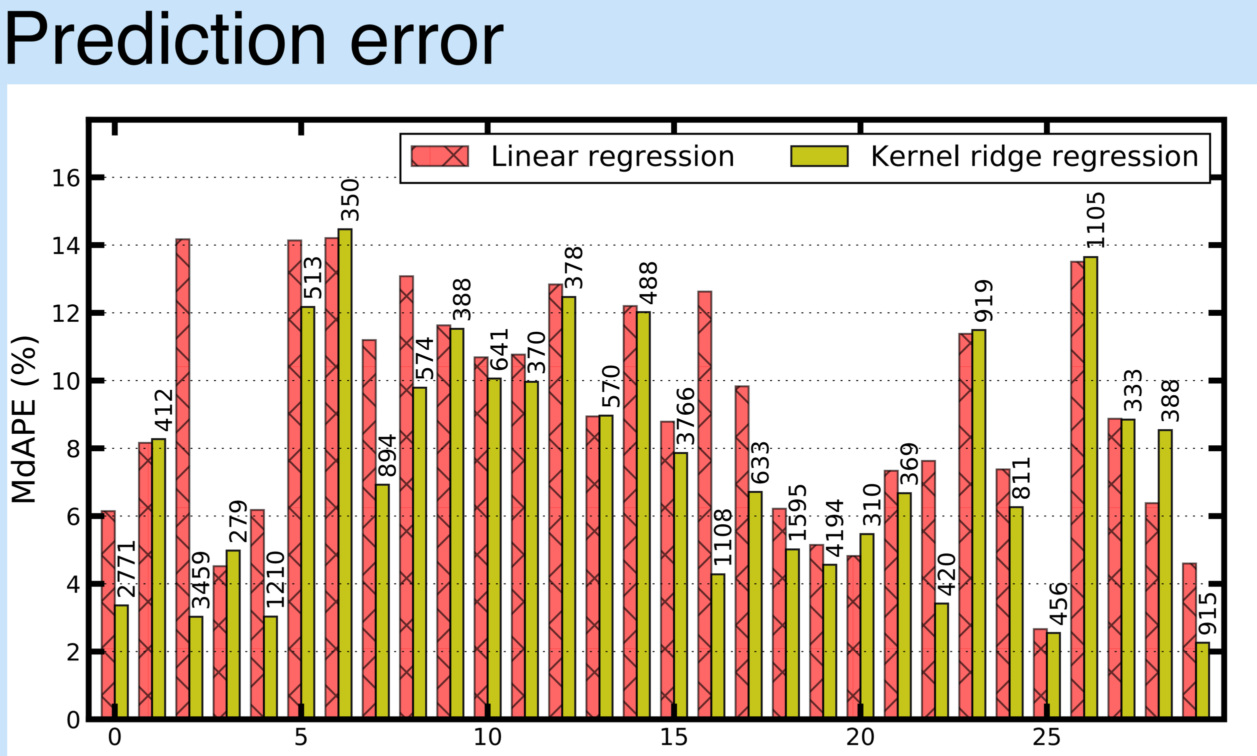


Tunable parameters matter



3. Prediction and model based feature importance

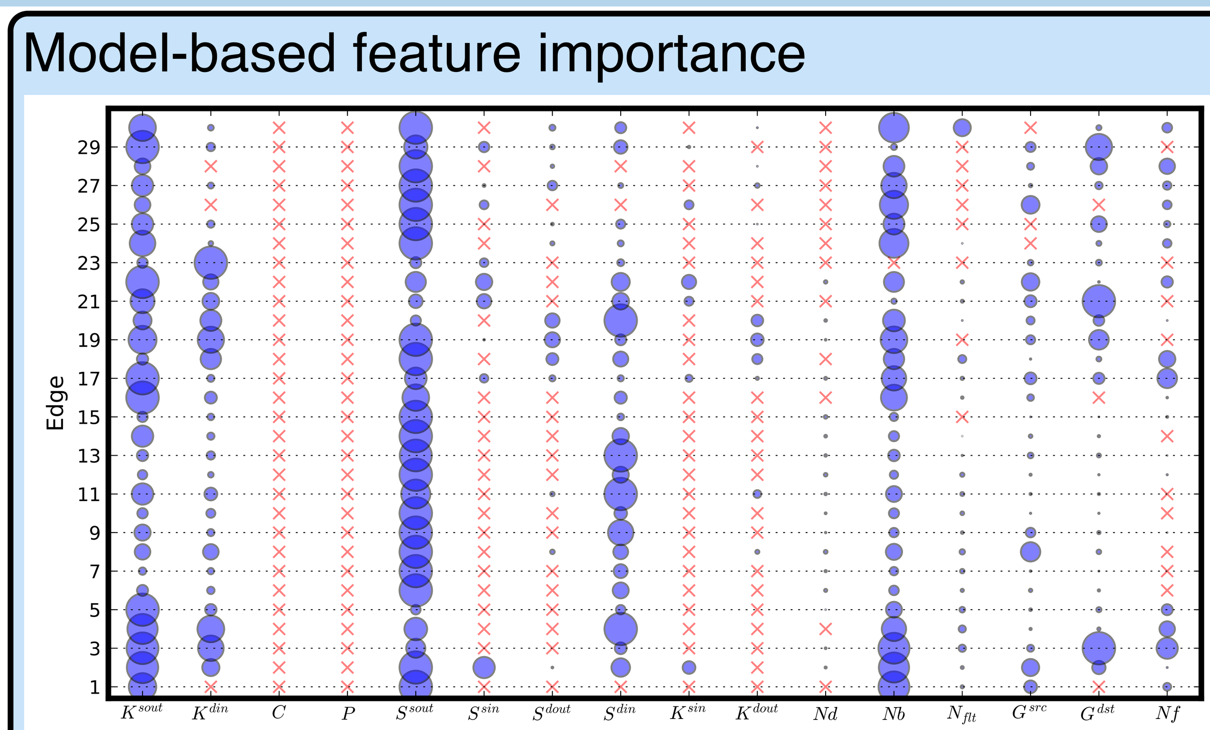
Prediction error



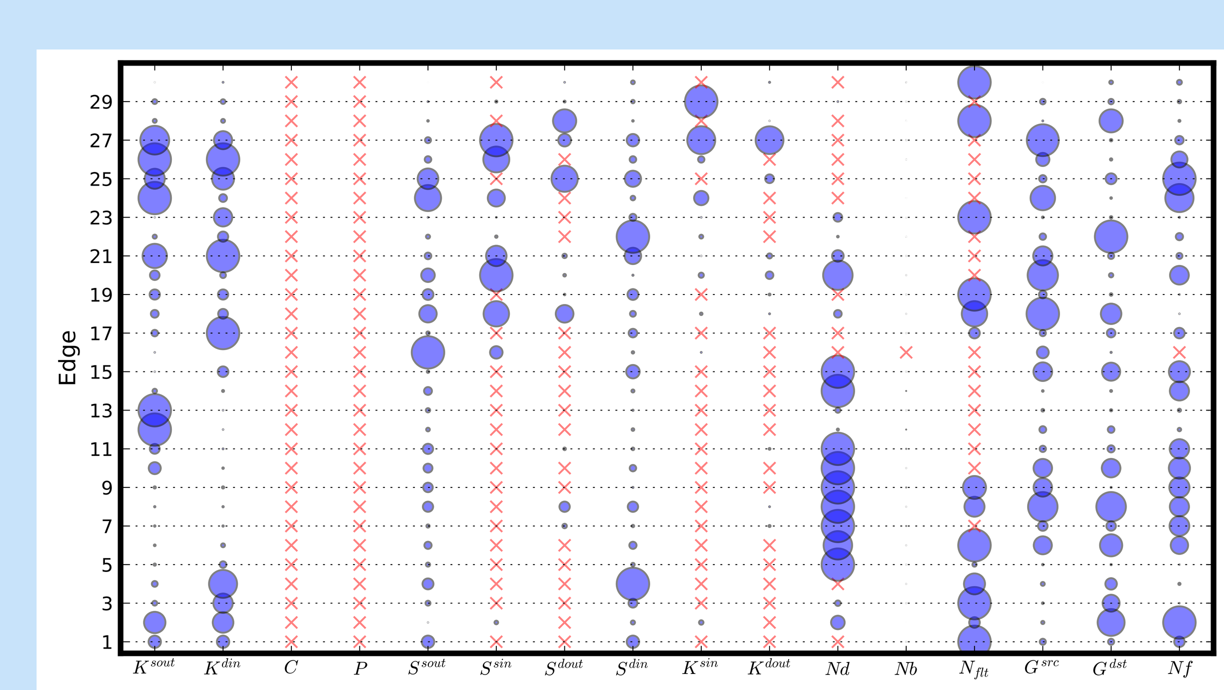
Acknowledgements

This material was supported in part by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. We thank Nagi Rao for useful discussions, Brigitte Raumann for help with Globus log analysis, Glenn Lockwood for help with experiments at NERSC, and the Globus team for much good work and advice.

Model-based feature importance



Linear regression model based feature significance



Nonlinear regression model (XGBoost) based feature significance

Circle size indicates the relative significance of features in the linear model, for each of 30 edges. A red cross means that the corresponding feature is eliminated because of low variance.