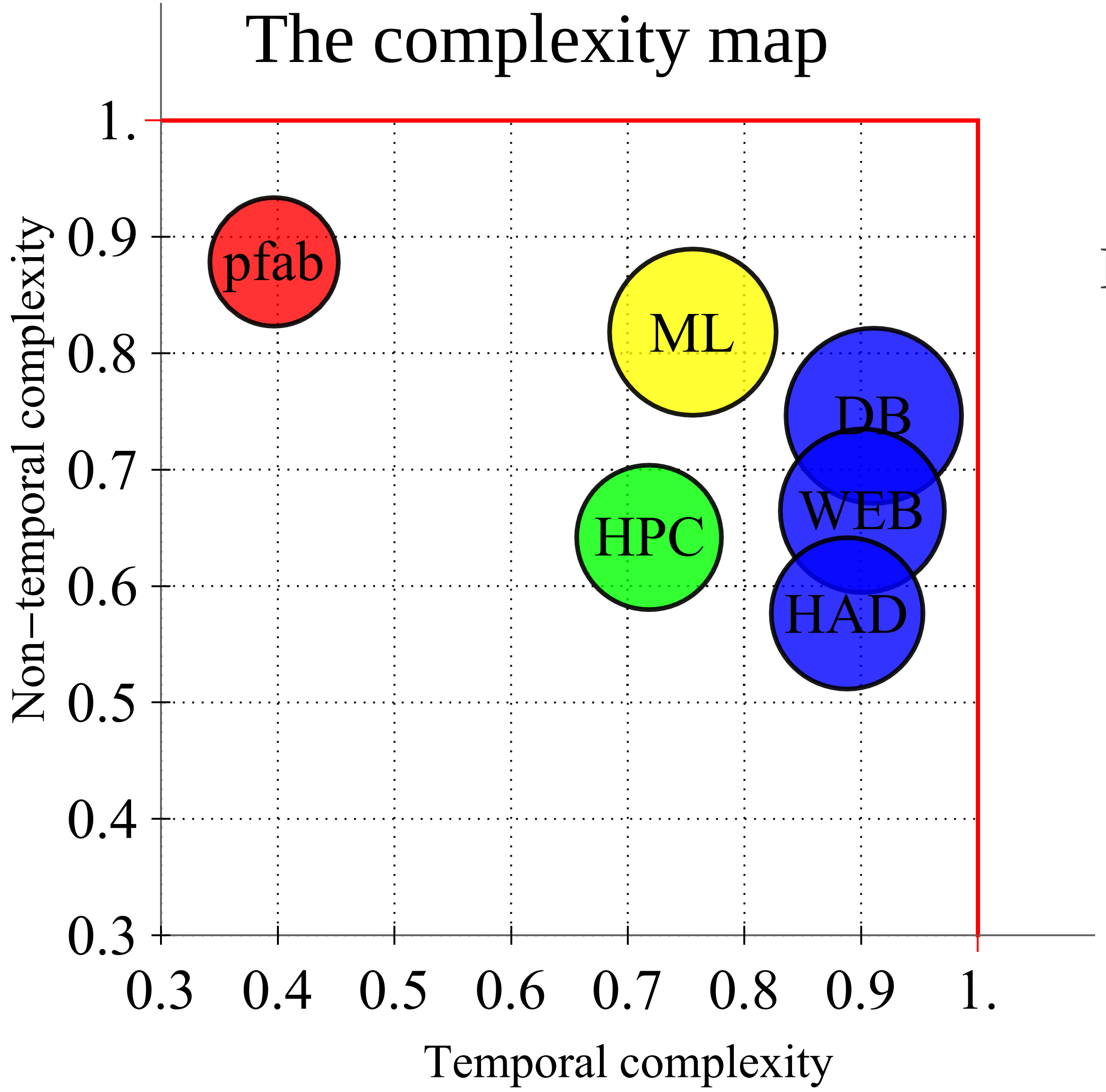
On the Complexity of Traffic Traces and Implications

CHEN GRINER

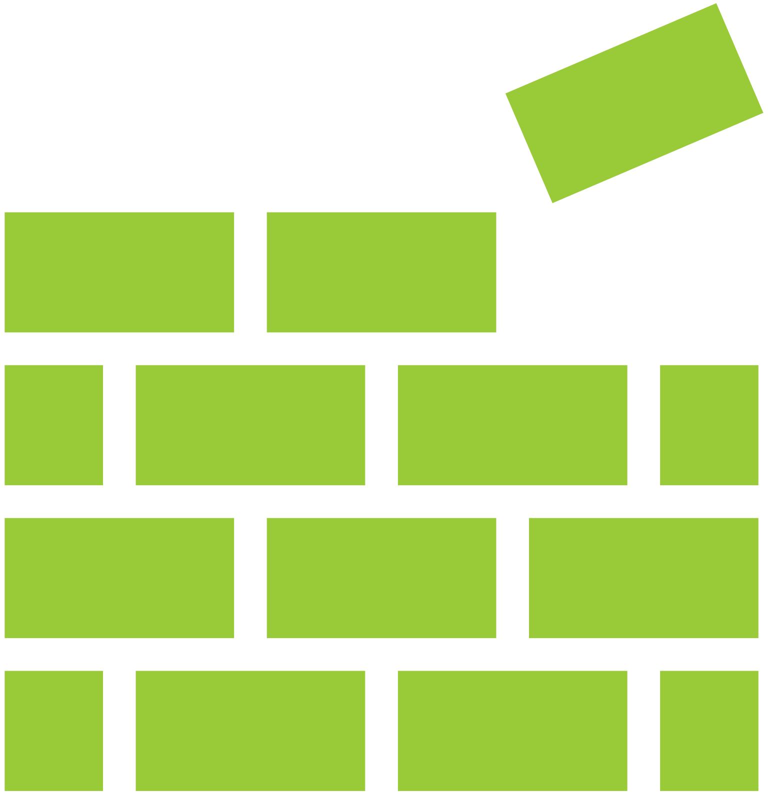
CHEN AVIN, MANYA GHOBADI, CHEN GRINER, STEFAN SCHMID

"It is not the strongest of the species that survives, nor the most intelligent that survives. But the one that is most adaptable to change."
(Leon C. Megginson)



Mapping a Landscape of Network Traffic

- Define
- Categorize
- Quantify
- *Map different types of structure*
- Better design of networks?

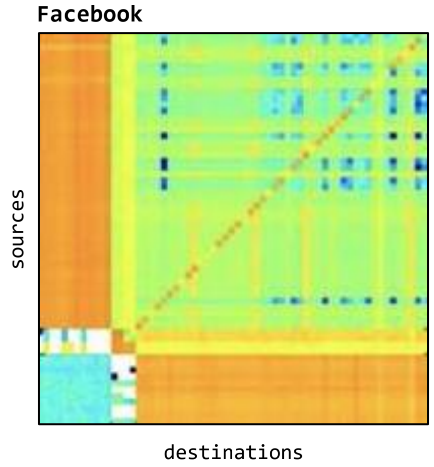


What is Structure?

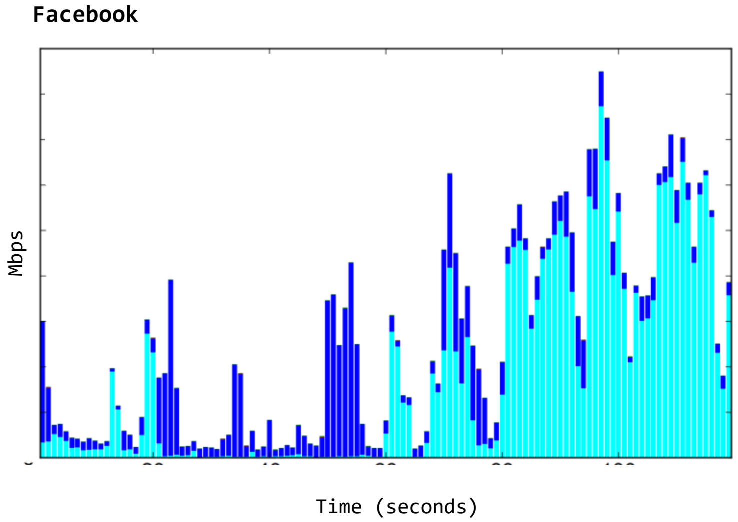
Visible Structure?

- Can we see patterns in real network traffic?
  - *Yes!*

Real

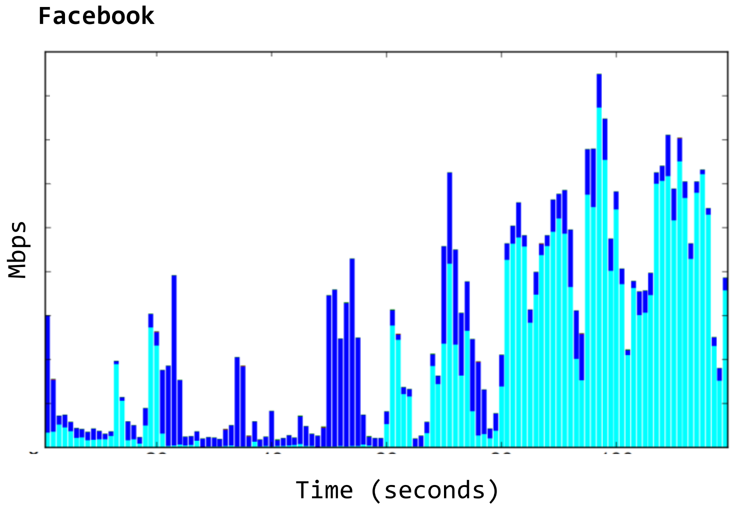


Rack-to-rack, Frontend cluster
FB @ SIGCOMM 2015

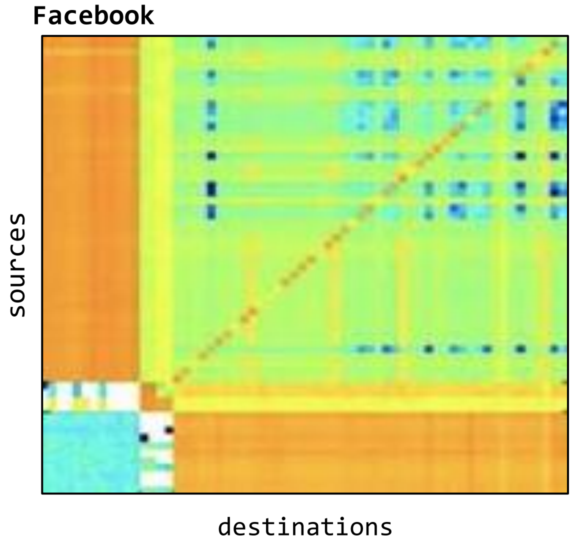


Structure: Two Types

- Temporal

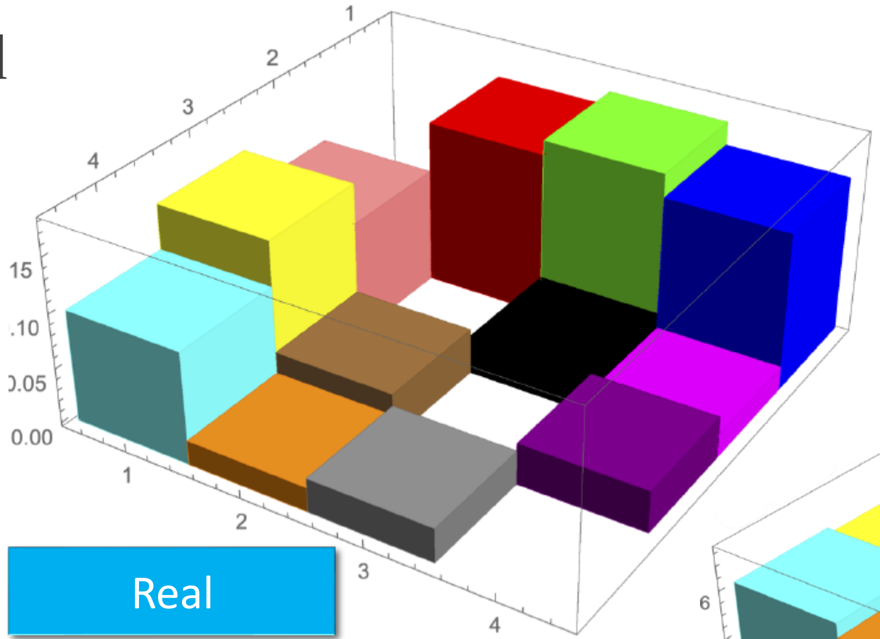


- Non temporal(spatial)

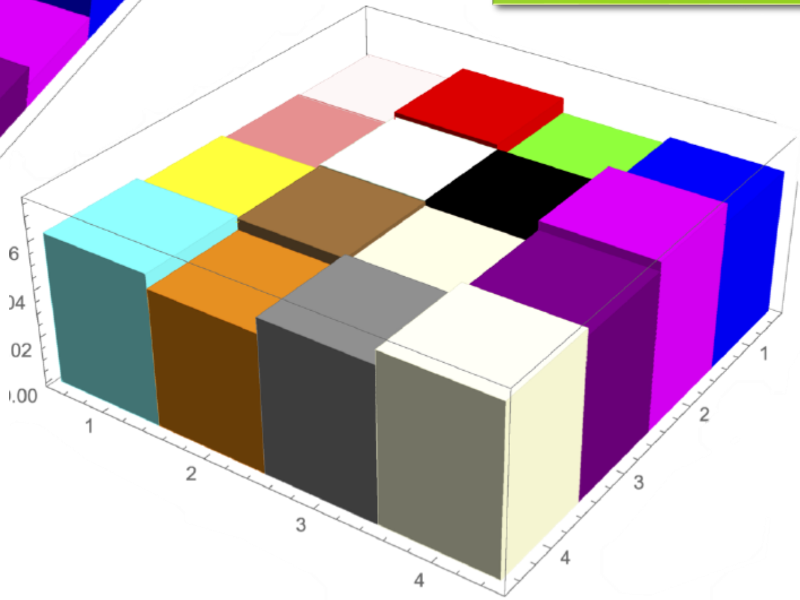


ML Example: Non temporal structure

- A traffic matrix of a distributed ML application
- Source destination denoted by color
- Height denotes amount of traffic
- **How would a matrix without structure look like?**



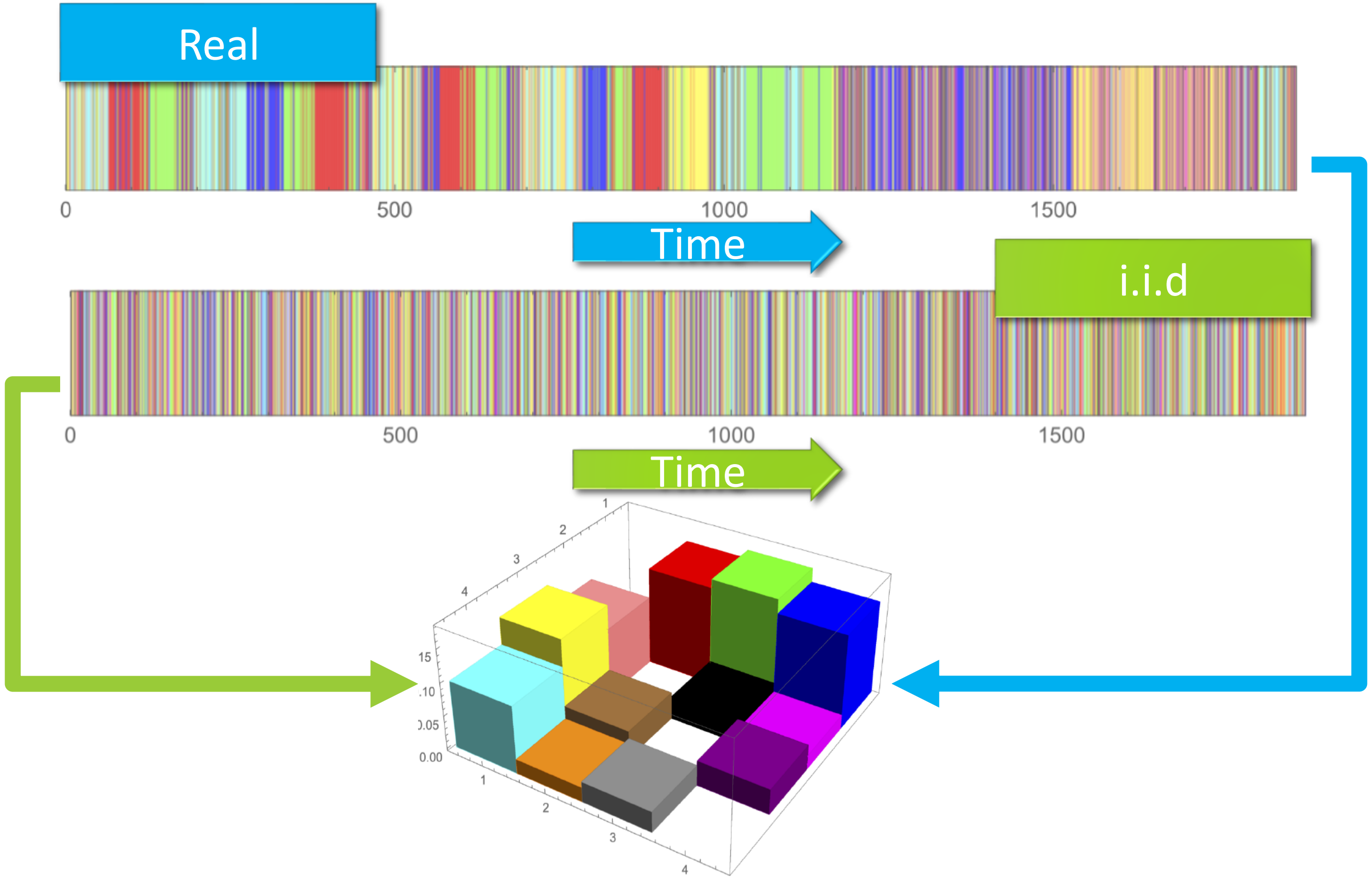
Real

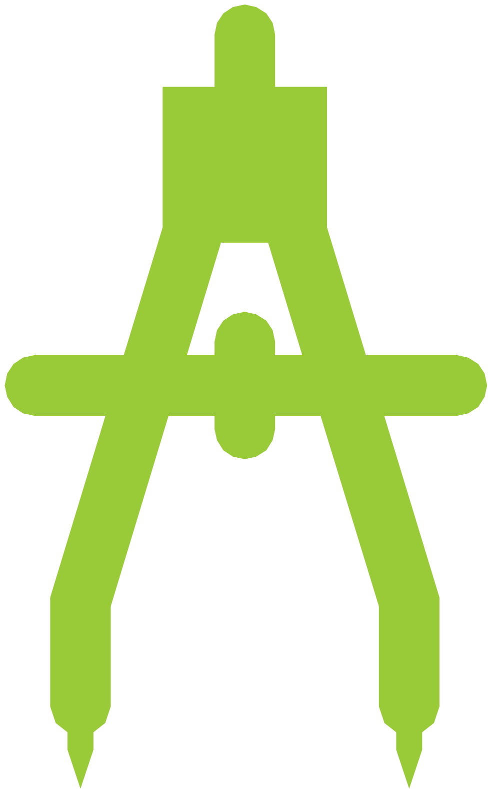


Uniform

Temporal structure

- Not all structure is related to frequency
- ***Temporal*** structure, represents the dependency of future events on recent events
  - Example: bursts of traffic
  - Both traces have the same traffic matrix but are different in time

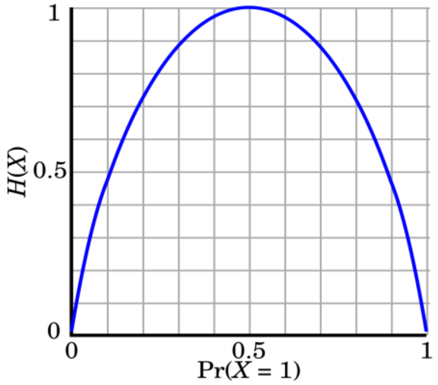




How to Measure Structure

An information theoretic perspective

- In Information theory *entropy* (entropy rate) is a measure "randomness"
- Lower *entropy* ⇒ More predictable traffic
- Traffic with more "structure" is less "complex"
- *Compression* offers a way to estimate *entropy*
- But how to represent traffic?



Traffic as a Network Trace

Source	Destination
(192.168.1.3)=s_0	(192.168.1.1)=d_0
s_1	d_1
s_2	d_2
s_3	d_3
s_4	d_4
s_5	d_5
s_6	d_6

Time

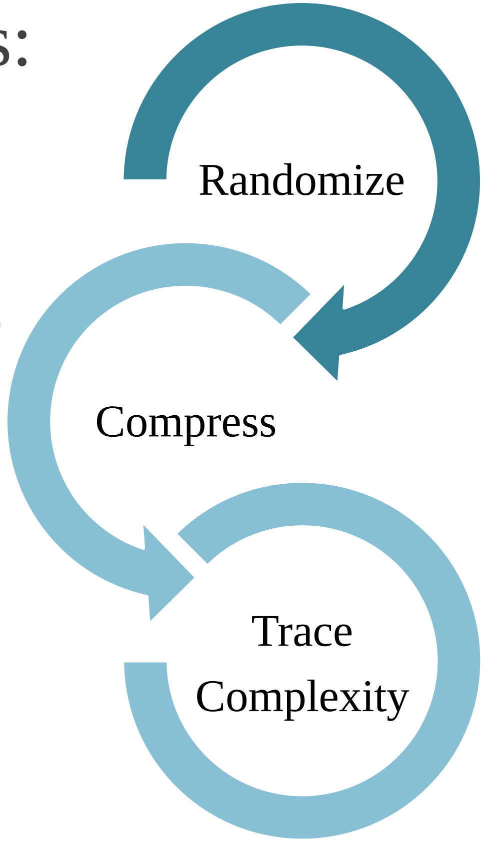
- A Simplified **time ordered** list of **source-destination** pairs
- How to use it?

Methodology

- Measuring complexity with two steps:

1. Sequentially randomize a trace, remove a specific types of structure

2. Compress the trace, compare their size.



Source	Destination
s_0	d_0
s_1	d_1
s_2	d_2
s_3	d_3
s_4	d_4
s_5	d_5
s_6	d_6

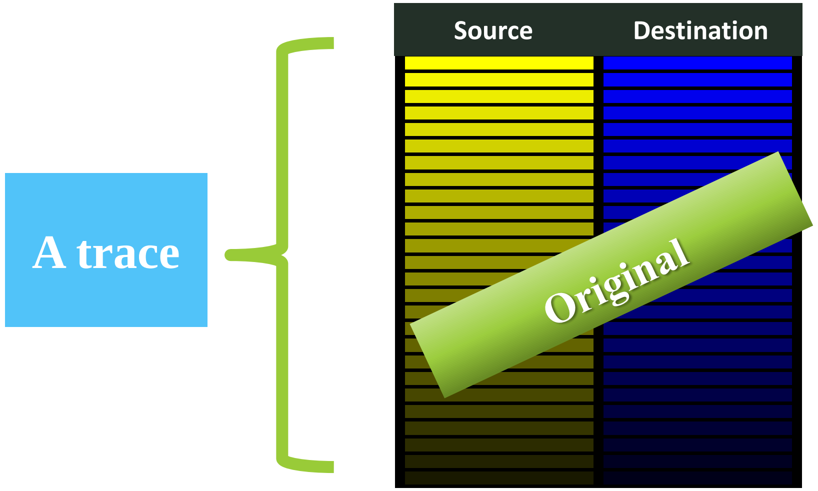
Systematic randomization

A trace

Source	Destination
s_0	d_0
s_1	d_1
s_2	d_2
s_3	d_3
s_4	d_4
s_5	d_5
s_6	d_6

Systematic randomization

Increasing complexity

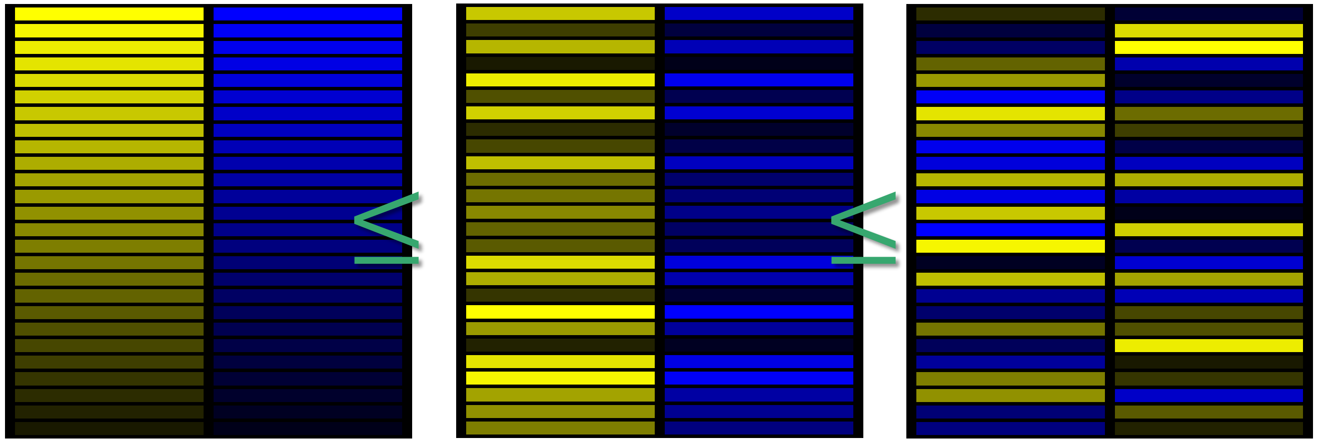


Original trace $\boldsymbol{\sigma}$
- ✓ Temporal
- ✓ Non Temporal

Row Randomized $\Gamma(\sigma)$
- X Temporal
- ✓ Non Temporal

Uniform Trace $\mathsf{u}(\sigma)$
- X Temporal
- X Non Temporal

Compression



Formal Definitions of Trace Complexity

Let $c(\sigma)$ is the *size* of a compressed trace σ:

We define:

Total complexity: $\psi(\sigma) = \frac{\mathrm{c}(\sigma)}{\mathrm{c}(\mathrm{u}(\sigma))}$

Uniform transformation

Temporal complexity:$\mathrm{T}(\sigma) = \frac{\mathrm{c}(\sigma)}{\mathrm{c}(\Gamma(\sigma))}$

Row transformation

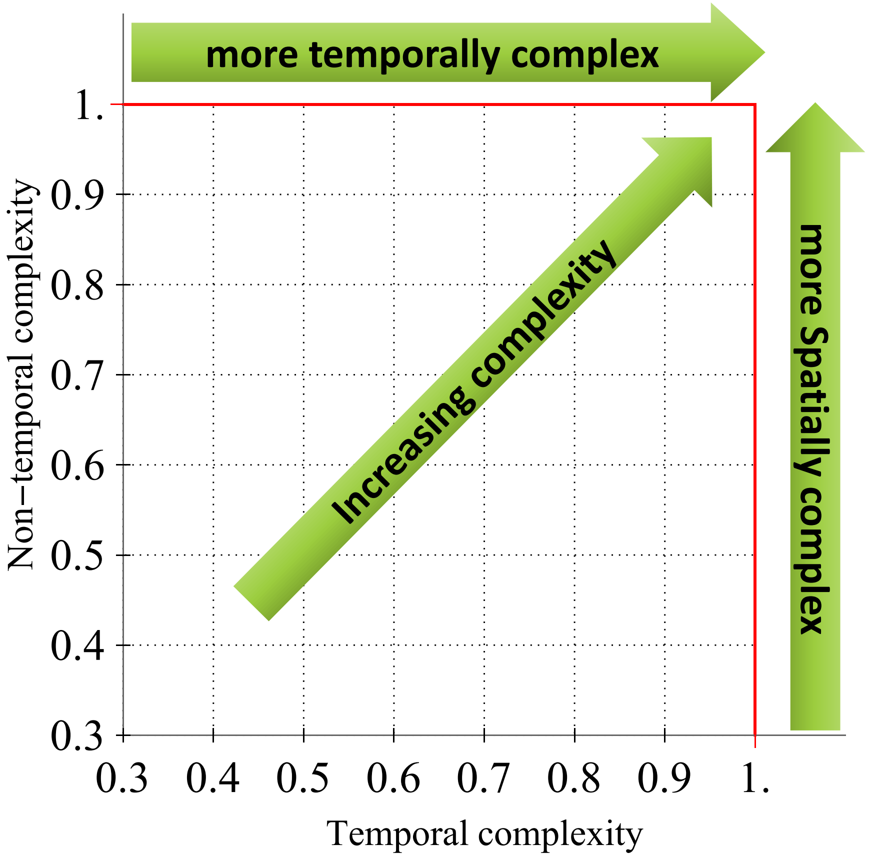
Non-temporal complexity: $\mathrm{NT}(\sigma) = \frac{\mathrm{c}(\Gamma(\sigma))}{\mathrm{c}(\mathrm{u}(\sigma))}$

Complexity is in the range of $[0 \ldots 1]$

Mapping Trace Complexity

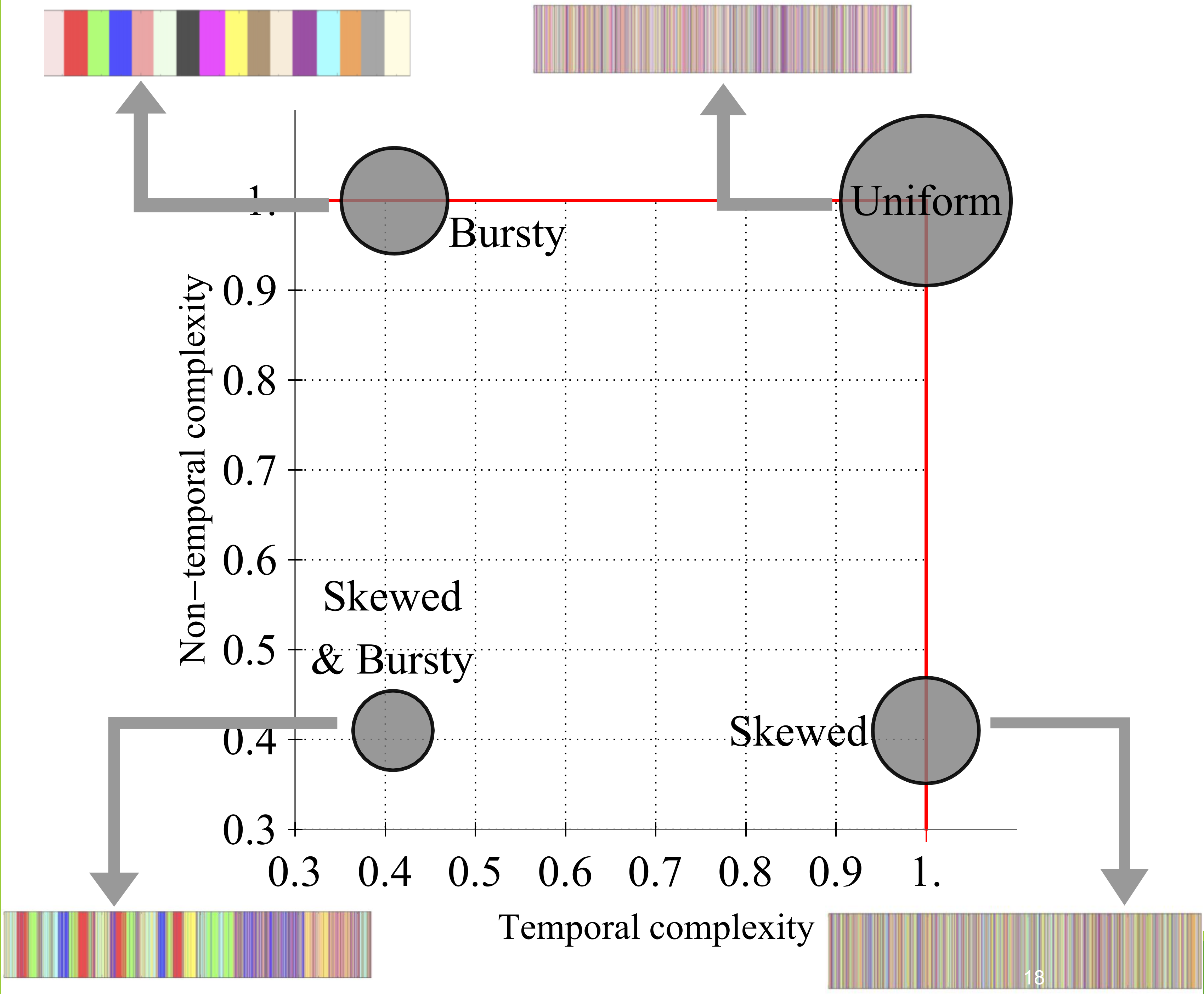
The Complexity Map

- *X* axis: temporal complexity
- *Y* axis: non-temporal complexity
- Rule of thumb:
  Closer to the axis's origin, means lower complexity



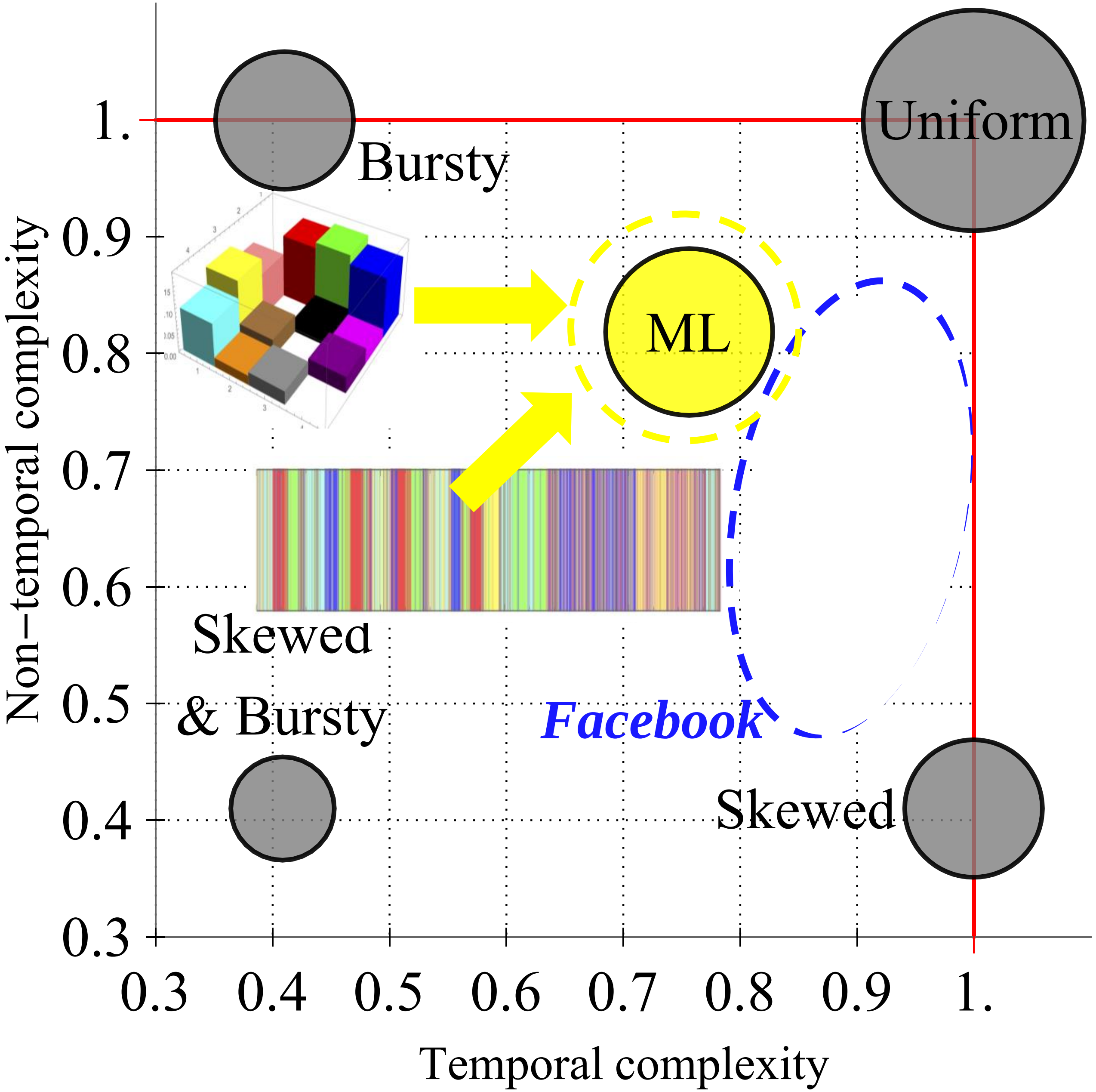
The Complexity Map

- Uniform Traffic
  - Lacking any structure, maximal entropy
- Bursty Traffic
  - Has temporal correlations
- Skewed Traffic
  - From a skewed distribution
- Skewed & Bursty Traffic
  - Has both temporal and non temporal elements



Case Study

- ML
- Facebook
  - Database, Web, Hadoop
- High Power Computing (HPC)
- pFabric



What can we expect to learn from trace complexity?

- Lower complexity means better optimization
- Identify and quantify different structures
- Compare different traces?
- Differentiate between different workloads?

Future work

- Test trace with more metadata
  - Interarrival times, ports etc
- What are the other dimensions of complexity?
- Practical implementation of complexity in online algorithms?
- Trace website:
  - https://self-adjusting.net
- Further details are found in the paper:

On the Complexity of Traffic Traces and Implications.
Chen Avin, Manya Ghobadi, Chen Griner, Stefan Schmid. ***Sigmetrics 2020***