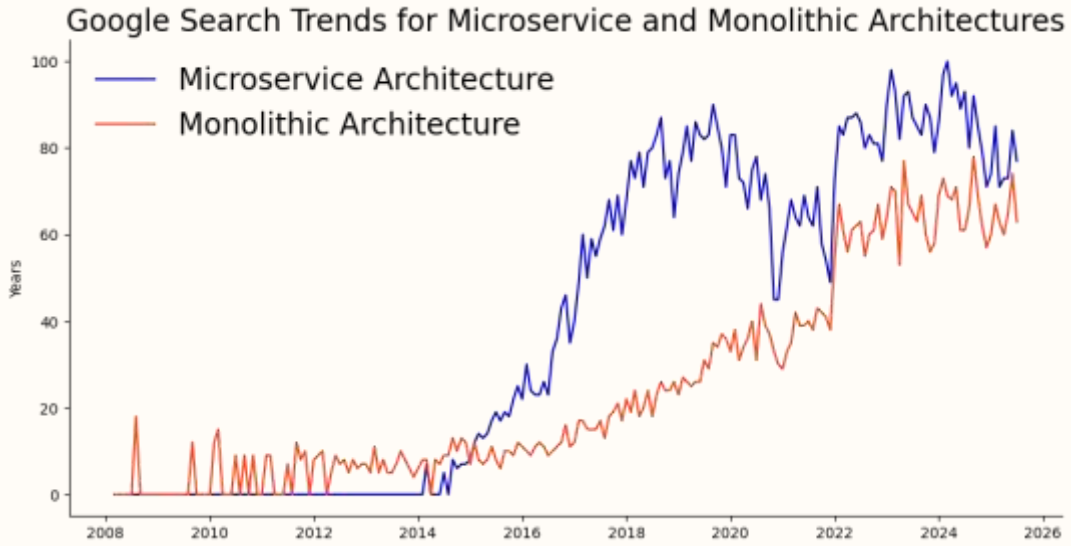
Complexity at Scale: Microservices at Alibaba

Giles Winchester

UNIVERSITY
OF SUSSEX

Why Microservices?

- They were a promise of a better way to develop applications:
  - Scalability
  - Fault isolation
  - Faster development
  - Improved organisational alignment

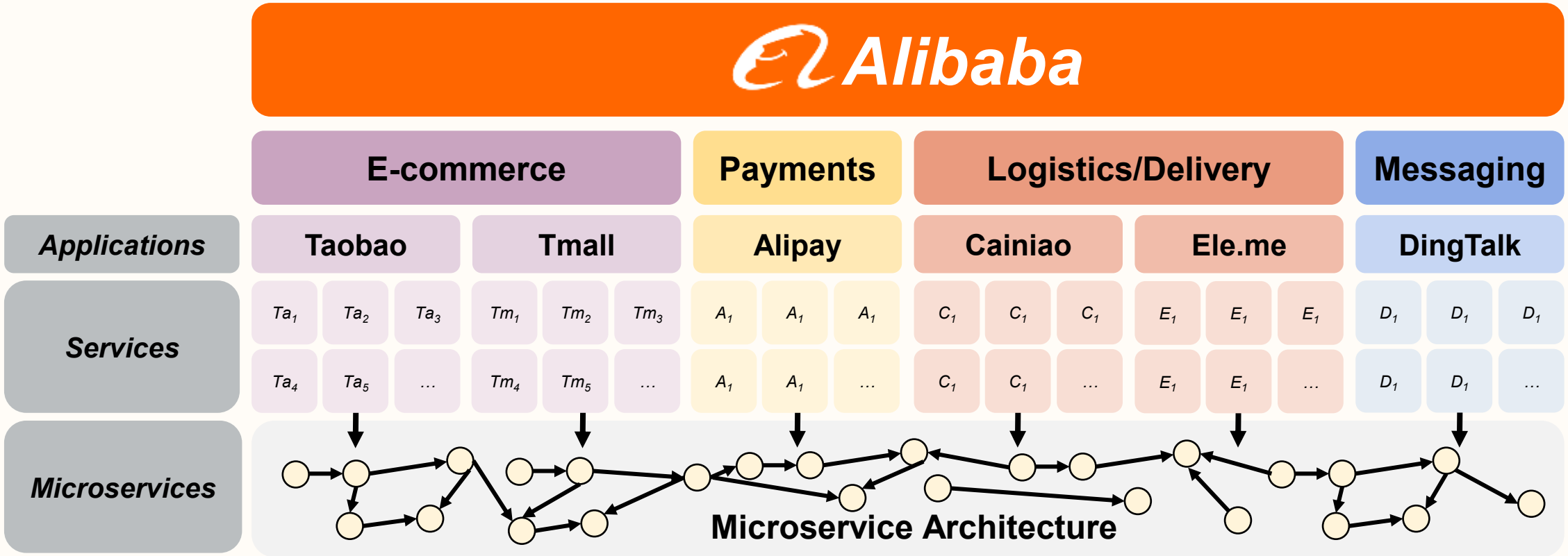


- "*the pains of microservices are mainly due to [their] intrinsic complexity*" [1]

UNIVERSITY OF SUSSEX

[1] "The pains and gains of microservices: A systematic grey literature review," *JSS*, vol. 146, pp. 215–232, Sep. 2018.

Microservices at Alibaba



UNIVERSITY OF SUSSEX

Scale: Raw Numbers

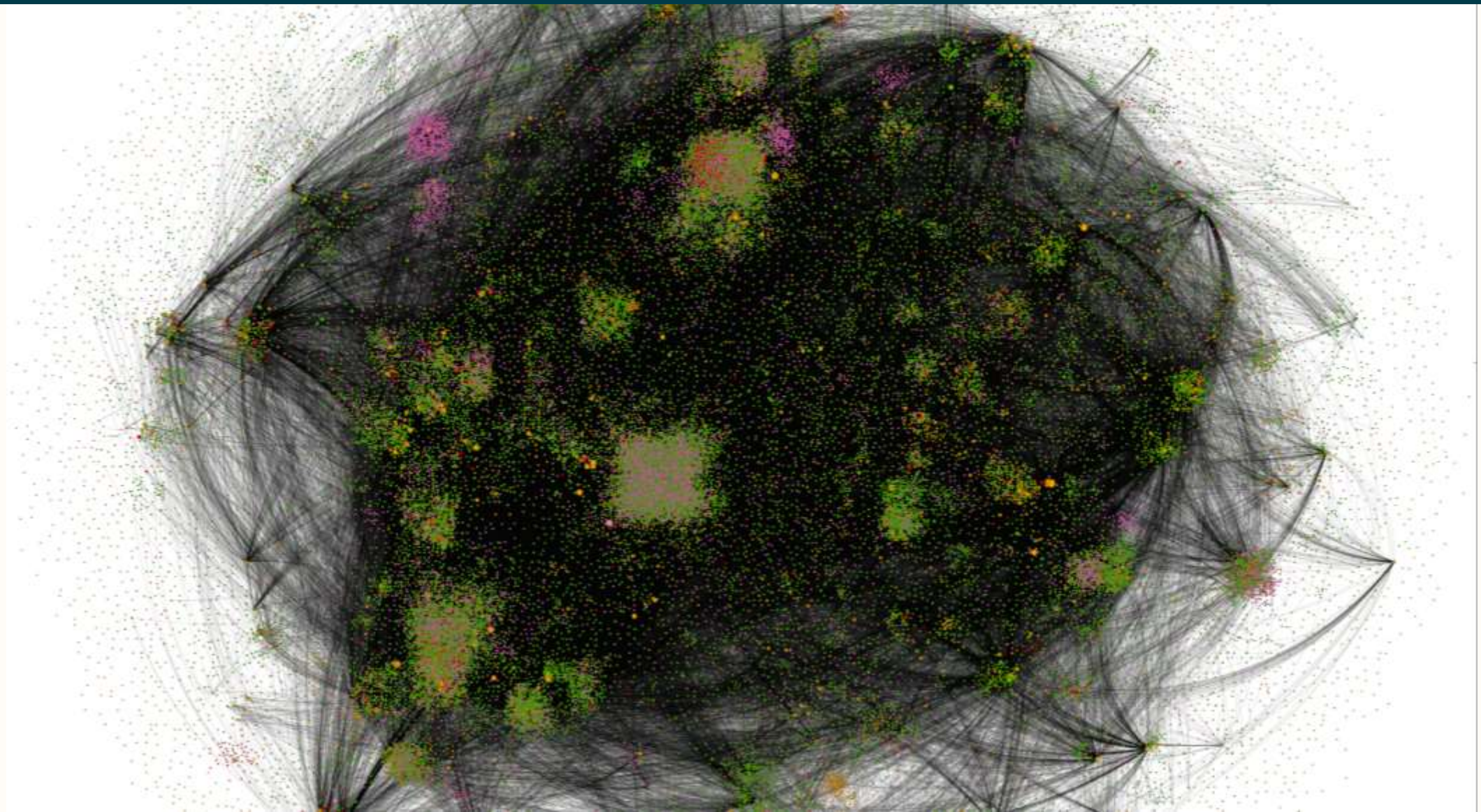
- **64,751** unique microservices
- **~1.8 million** microservice instances

- Supporting **~1 million** front-end service functionalities.

- Over 14 days received **15 billion** front-end requests which resulted in **97 billion** calls between microservices.
  - Sampled at 0.5% so actually ~300 billion and ~2 trillion front-end requests and calls

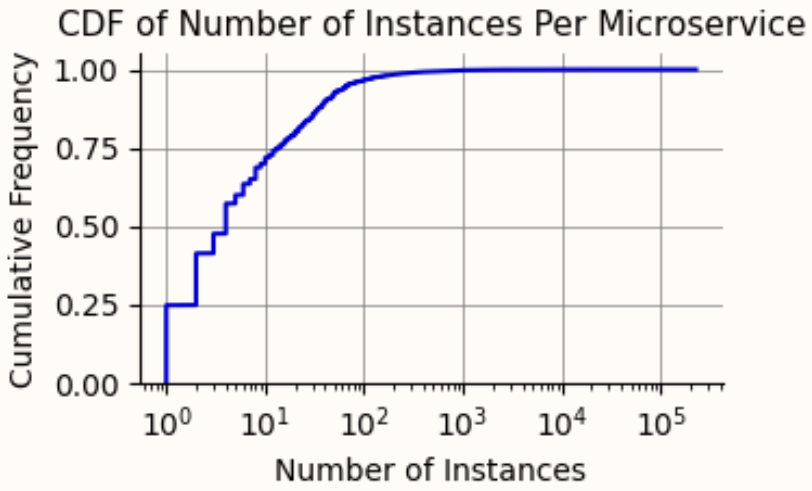
UNIVERSITY OF SUSSEX

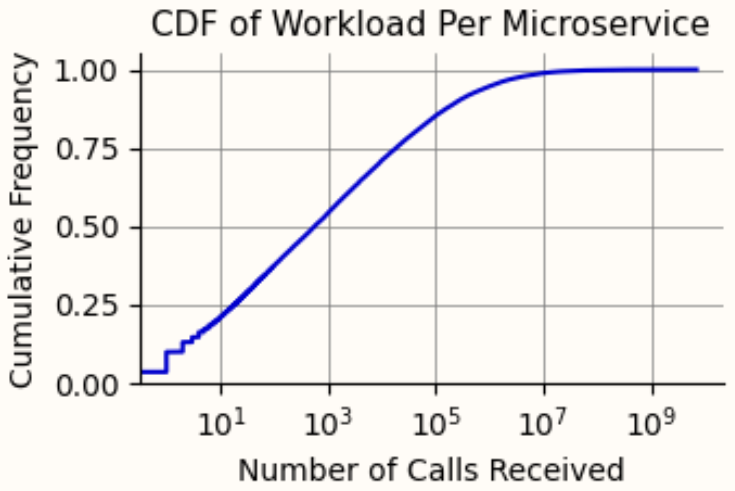
Scale: Dependency Topology



Heterogeneity: Microservice Inequality

- Microservices are **not** born equal; heavy-tailed distributions everywhere.
  - 1% of microservices account for **50%** of the deployed instances.
  - 1% of microservices account for **87%** of all workload.

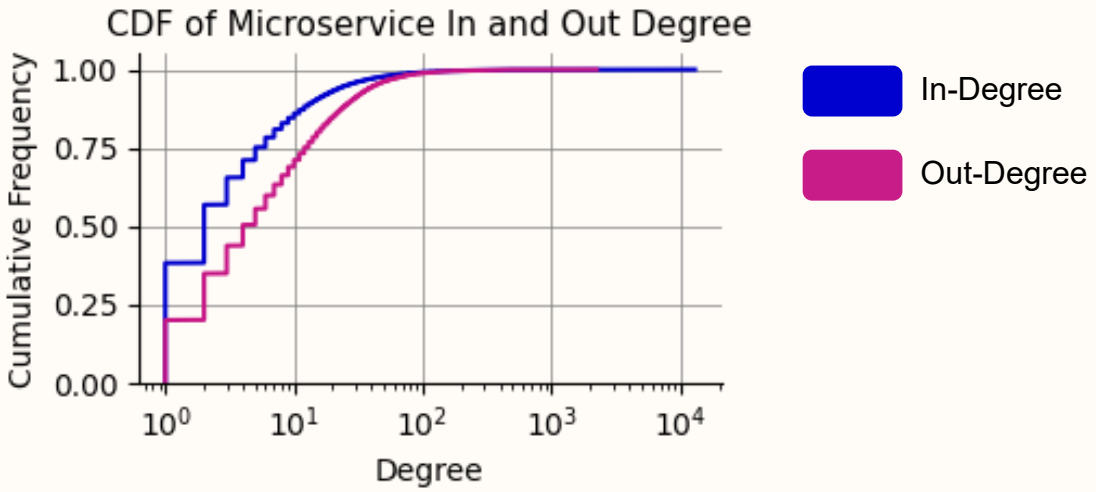




UNIVERSITY OF SUSSEX

Heterogeneity: Microservice Inequality

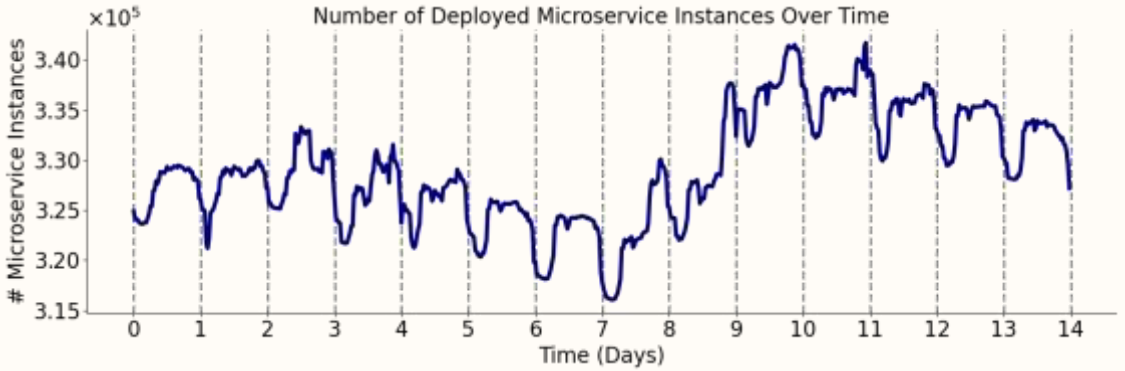
- Microservices are **not** born equal; heavy-tailed distributions everywhere.
  - Median dependency in-degree was **2** but maximum was **13,000**.
  - Median dependency out-degree was **4** but maximum was **2,200**.



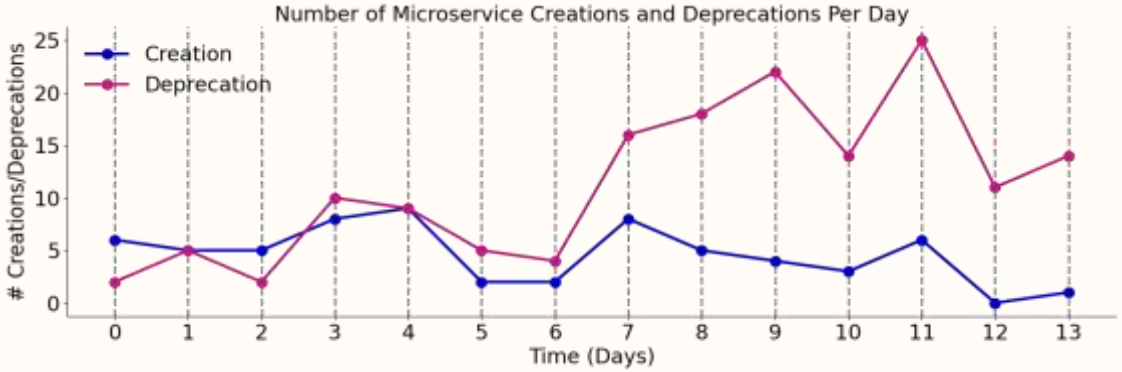
UNIVERSITY OF SUSSEX

Dynamicity: Evolutionary Architectures

- Microservice architectures are non-stationary
  - Horizontal scaling means that the number of deployed instances per microservice can change dramatically each day.



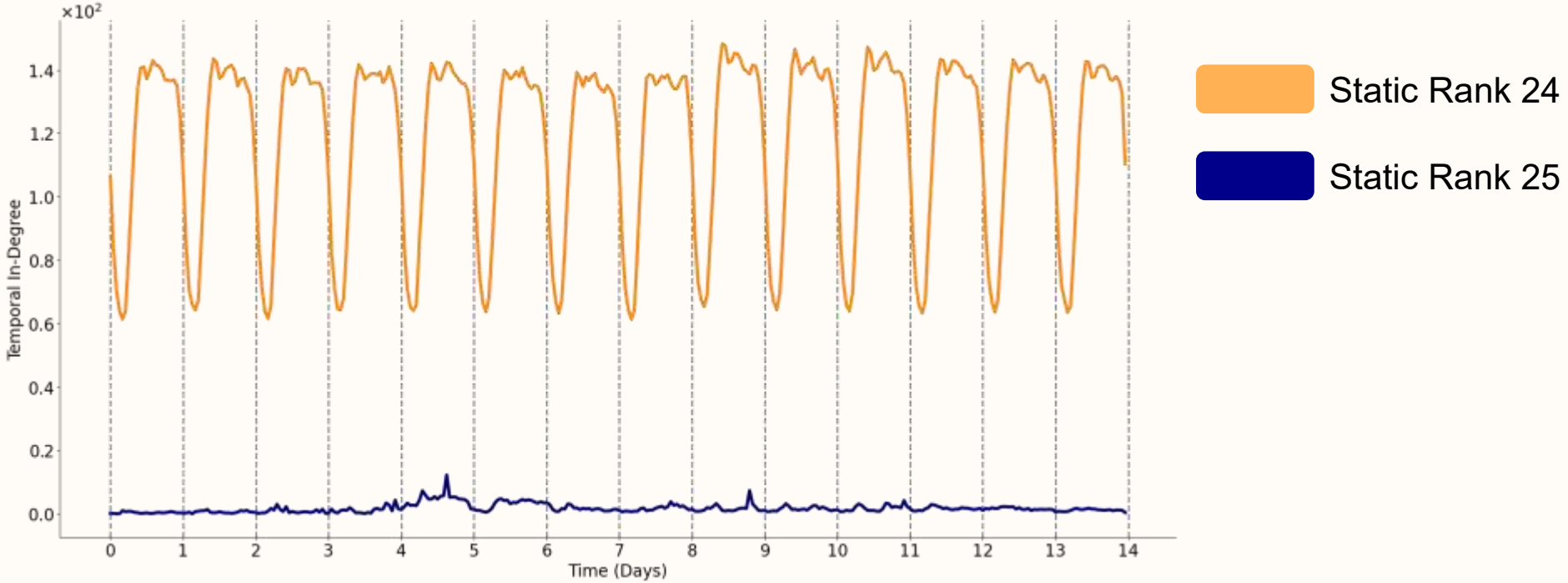
- Microservice architectures evolve daily.
  - **~20** microservices created and deprecated on average each day.



UNIVERSITY OF SUSSEX

Dynamicity: Time-Varying Dependencies

- Static views of dependencies do not necessarily align with temporal views at runtime.



UNIVERSITY OF SUSSEX

Thank you for listening – Q&A

G. Winchester et al, “Complexity at Scale: A quantitative analysis of an Alibaba microservice deployment,” *arXiv preprint arXiv:2504.13141*, 2025. Available: https://arxiv.org/abs/2504.13141

UNIVERSITY OF SUSSEX