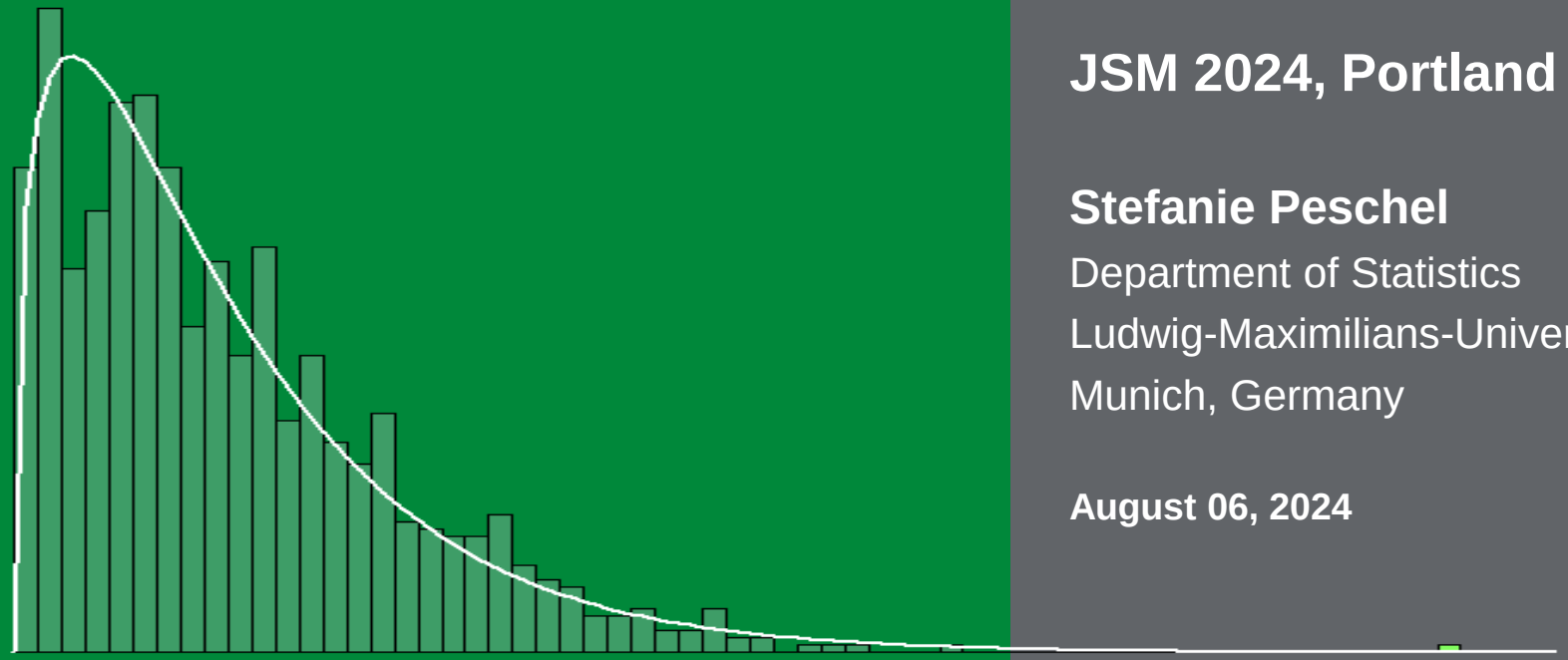
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FACULTY OF MATHEMATICS, INFORMATICS, AND STATISTICS

DEPARTMENT OF STATISTICS

Mind your zeros: accurate p-value approximation in permutation testing



Bio DS Lab

Biomedical Statistics & Data Science Lab

mcml Munich Center for Machine Learning

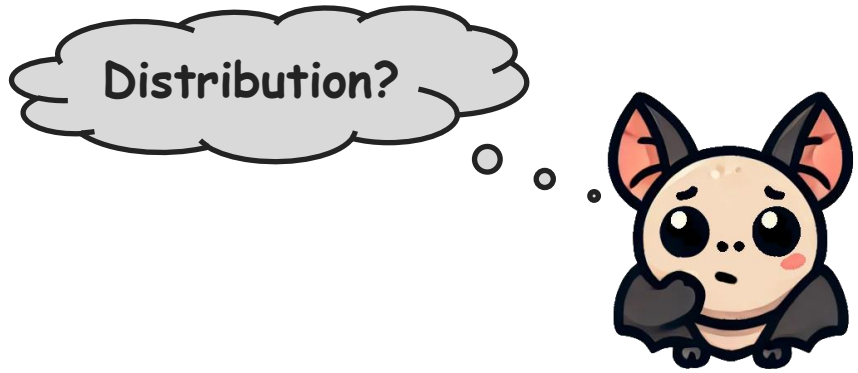
JSM 2024, Portland

Stefanie Peschel
Department of Statistics
Ludwig-Maximilians-Universität München (LMU)
Munich, Germany

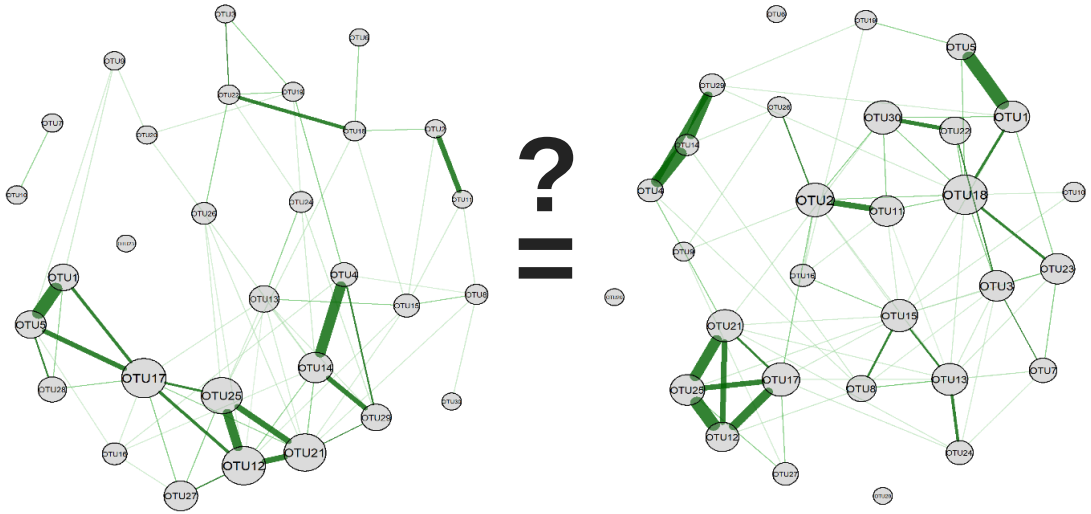
August 06, 2024

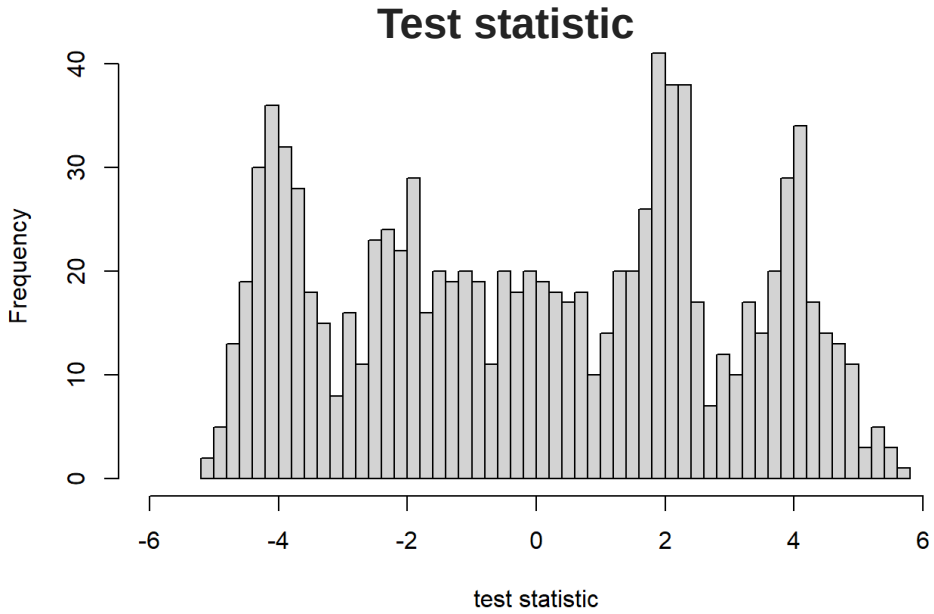
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Hypothesis testing



Different network properties?



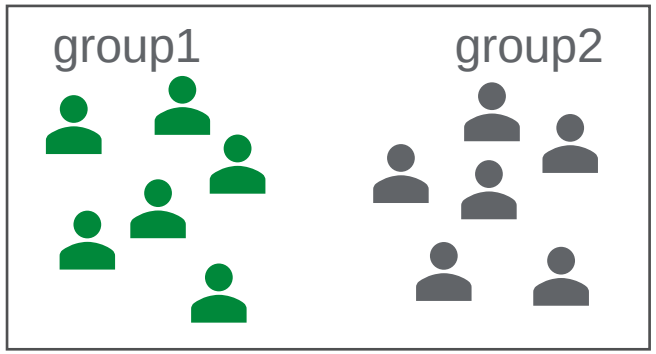


- No classical distribution
- Parametric assumptions not met

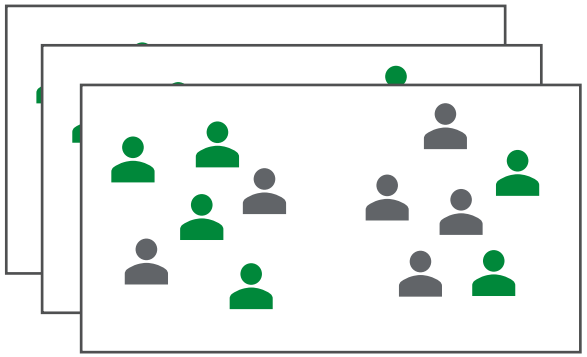
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Permutation testing

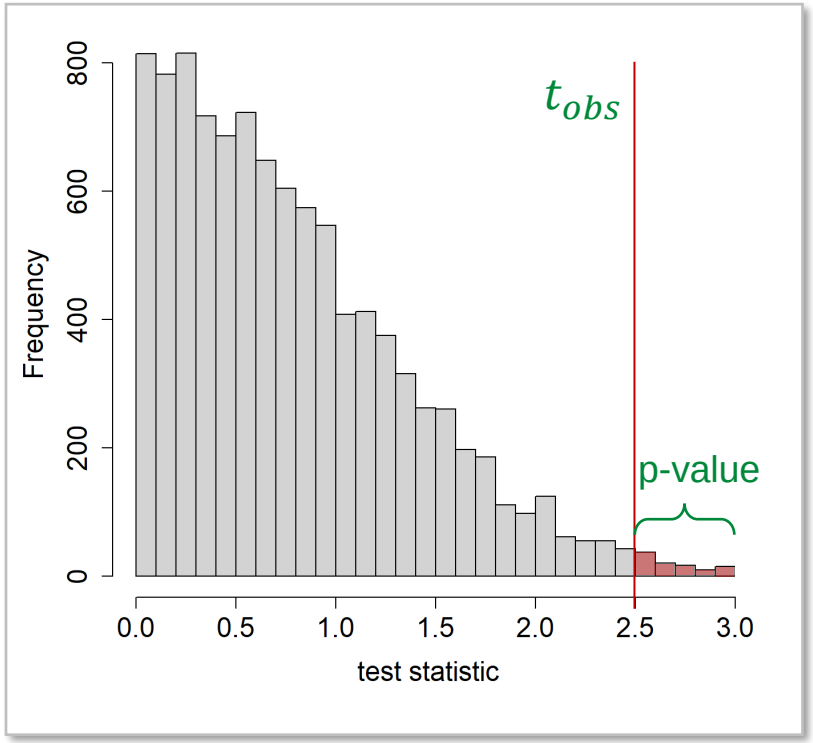
Observed data



Permutations



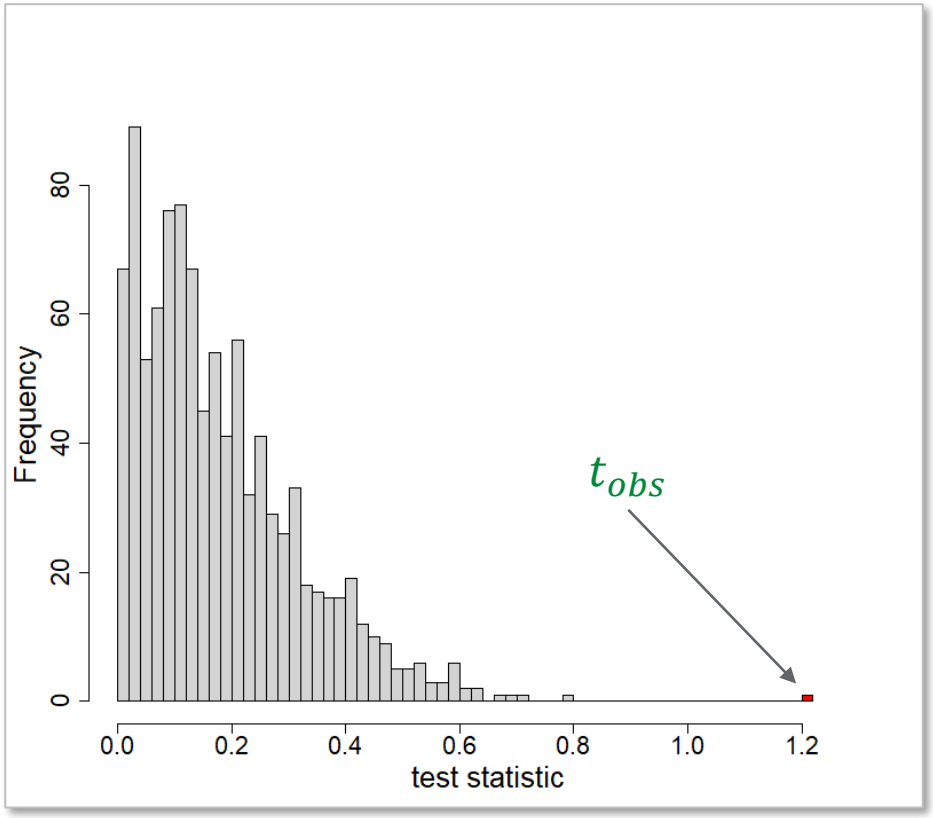
Distribution of t under H_0:



Empirical p-value:

$$p = \frac{\sum_{n=1}^{N} \boldsymbol{I}(t^* \geq t_{obs}) + 1}{N + 1}$$

Low number of permutations (here: N = 1000)



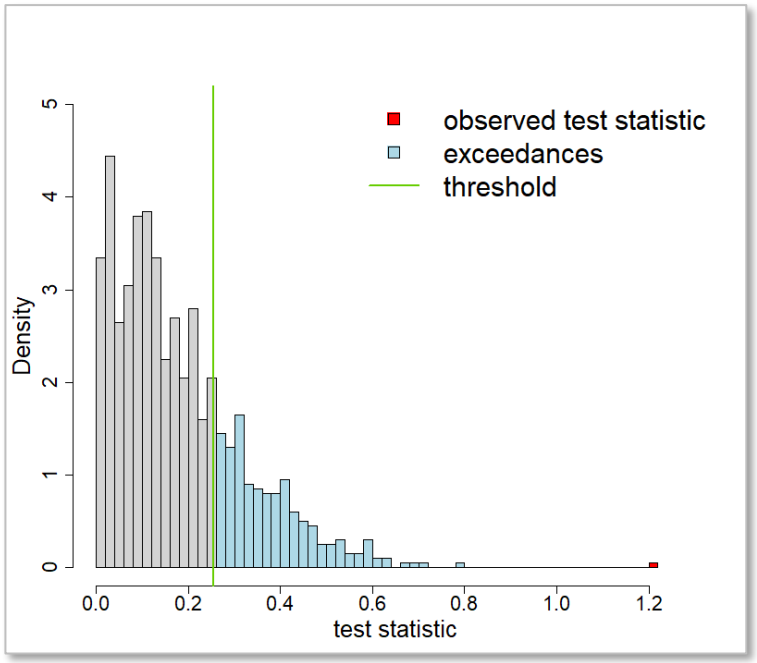
Empirical p-value:

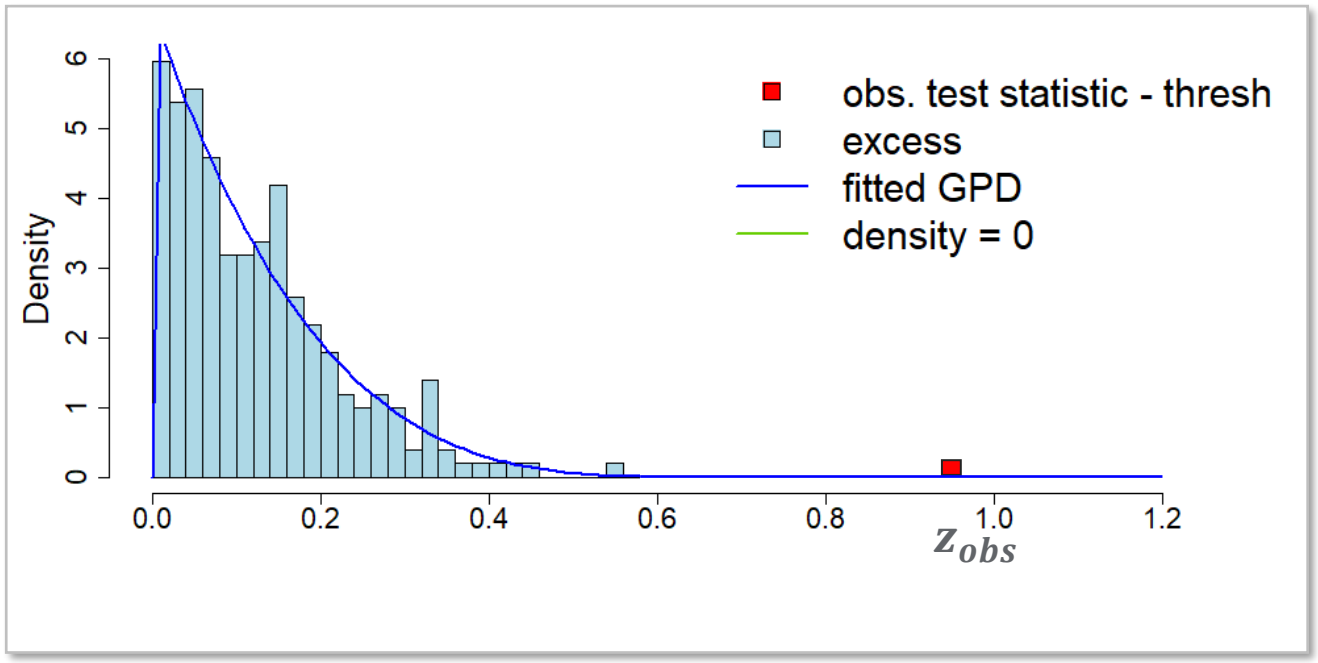
$$p = \frac{1}{1001} \approx 0.001$$

→ **Low power in multiple testing setting**

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

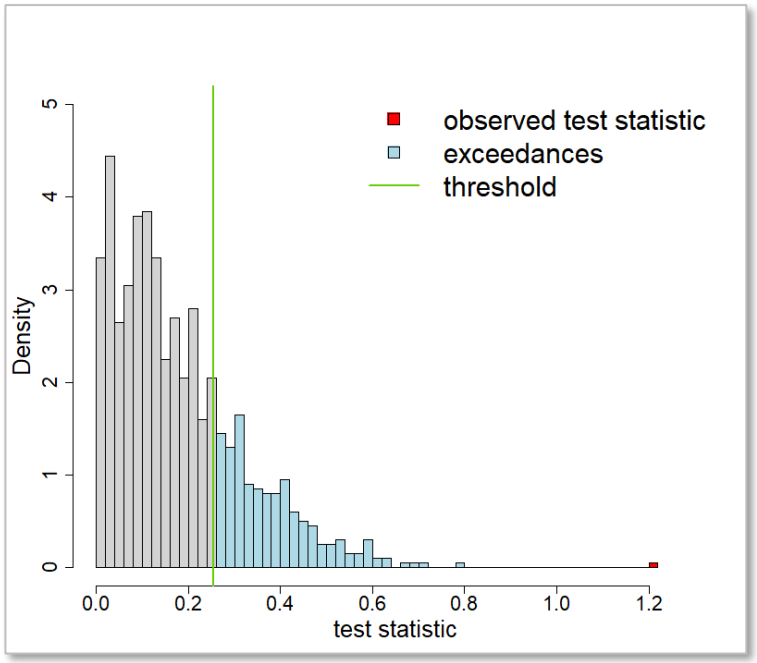
P-value approximation with the Generalized Pareto Distribution (GPD)

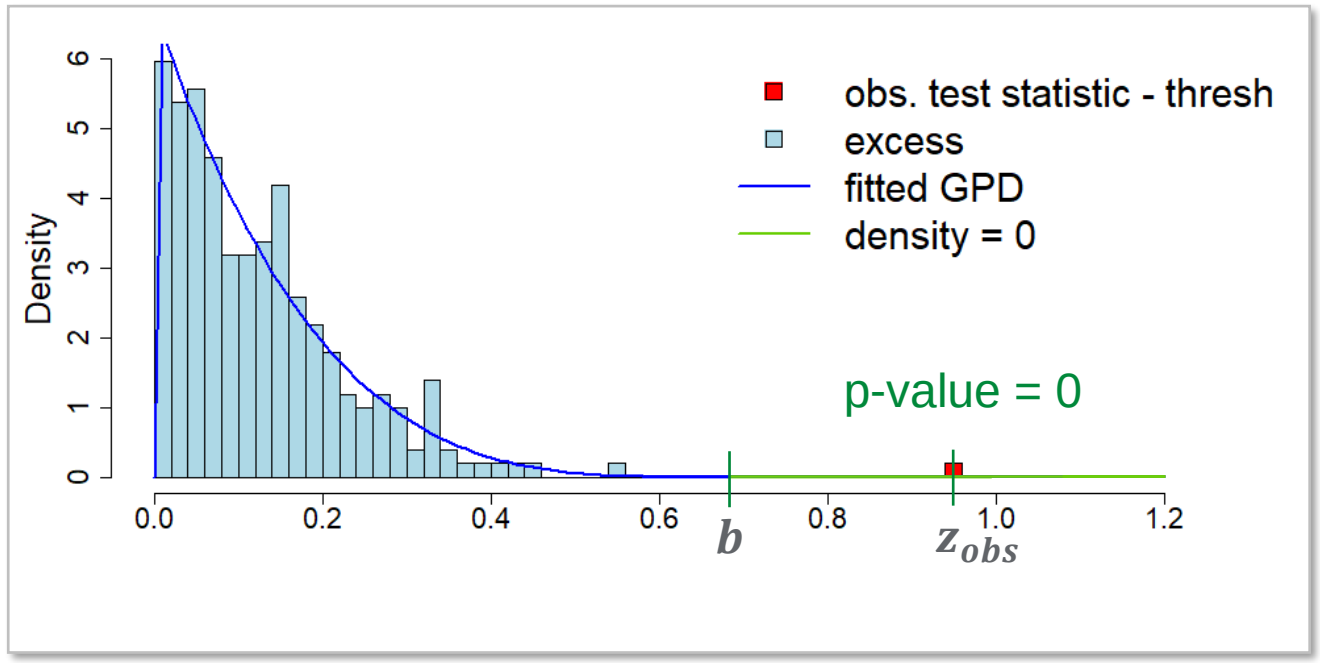




$$pval = \frac{N_{exc}}{N}\left(1 - F(t_{obs} - thresh)\right)$$

P-value approximation with the Generalized Pareto Distribution (GPD)





Estimated GPD parameters:

Shape: $\xi = -0.21785$ → Bounded support

Scale: $\sigma = 0.14890$

P-value is zero for extreme test statistics!

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Interested in our proposed approach?

Visit me in halls C&D

at poster number **11.**

mcml
Munich Center for Machine Learning

Get in touch:

Stefanie Peschel
LMU Munich, Germany

stefpeschel.de

github.com/stefpeschel

stefanie.peschel@stat.uni-muenchen.de

permAprox