University of Zurich UZH

SVA NATIONAL VETERINARY INSTITUTE

GILLES KRATZER, APPLIED STATISTICS GROUP, UZH

ARIANNA COMIN, DEP. OF DISEASE CONTROL AND EPIDEMIOLOGY, SVA

SVEPM WORKSHOP, UTRECHT 27.03.2019

ADVANCED BAYESIAN NETWORK MODELLING

Outline of the talk

‣ Mixed models – correction for grouped data

‣ Heuristic search

‣ MCMC over structures

‣ *Other advanced methods/features:*

  ‣ *Scoring system*

  ‣ *Tunable parameter prior*

  ‣ *Structural prior*

  ‣ *Data separation*

  ‣ *Covariate adjustment*

  ‣ *Likelihood contribution*

Correction for grouped data

‣ The way the data were collected has a clear **grouping aspect**
‣ Then potential for **non-independence** between data points
‣ Lead to analyses which are **over-optimistic**
‣ As the **true level of variation** in the data is **under-estimated**
‣ Could impact study result … or not!
‣ **Good practice to check!**

In practice:
‣ Random effect
‣ **GLM** -> **GLMM** for each node
‣ Fit the DAG and check the posterior distribution (**widening**)
‣ If needed one can **incorporate random effect** in the **scoring** scheme

Pitfalls:
‣ Hugh computational complexity!

Find maximum a posteriori score
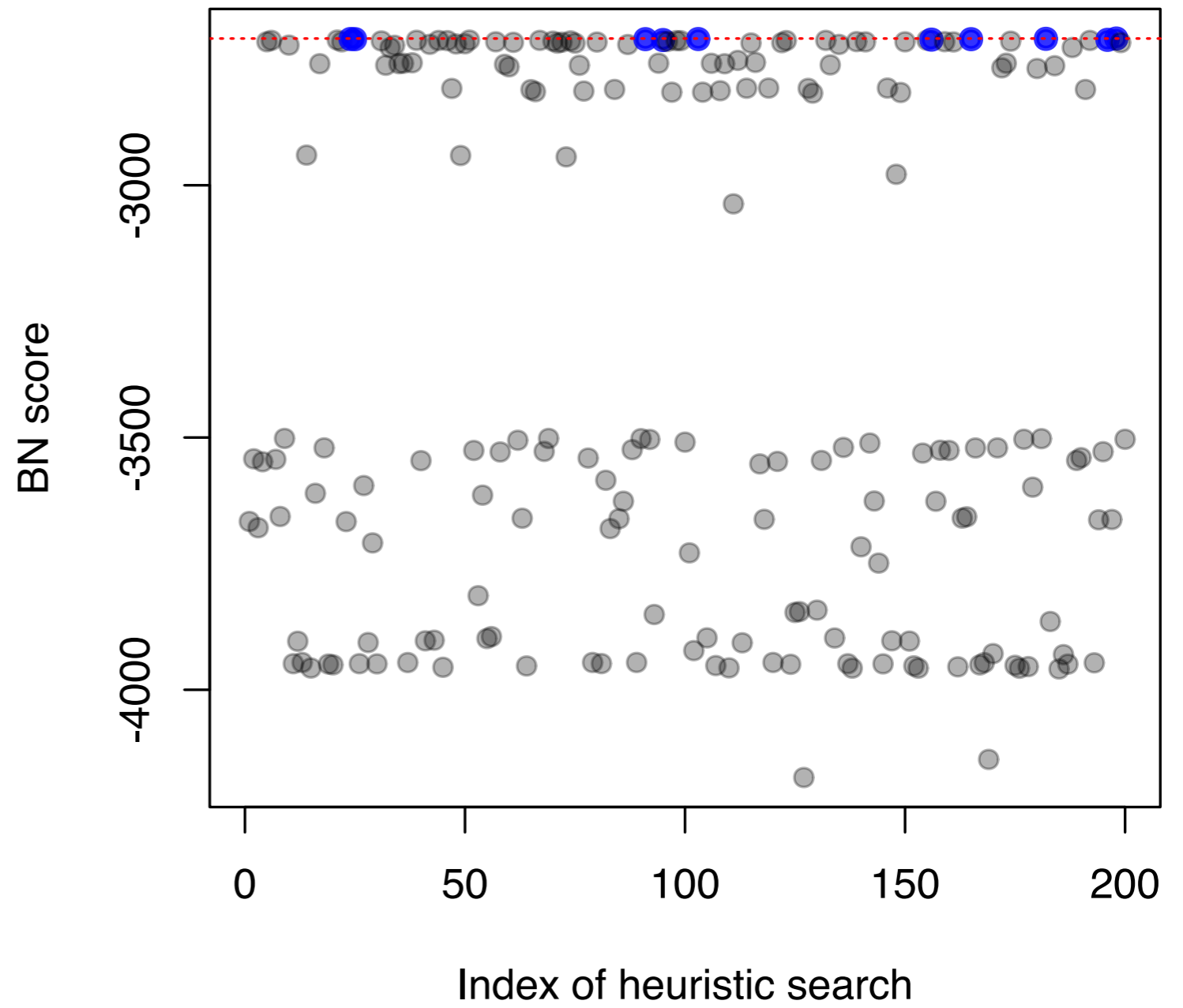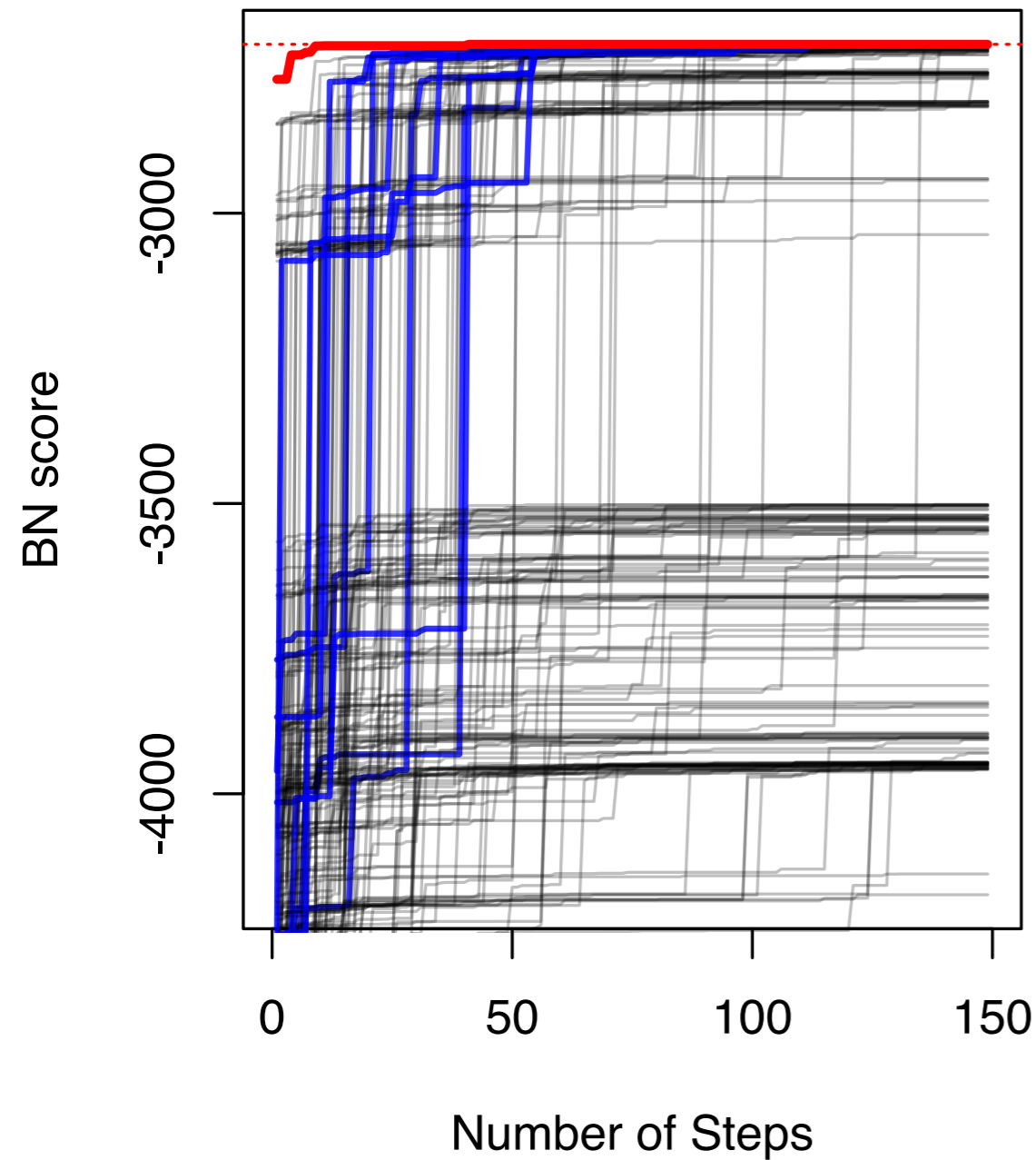- ✓ Exact search
- ‣ Heuristic search
- ‣ MCMC over structures

Heuristic search: Greedy Hill-Climbing

‣ Simplest heuristic local search
  ‣ Start with a given network
    ‣ empty network
    ‣ best tree
    ‣ a random network
  ‣ At each iteration
    ‣ Evaluate all possible changes
    ‣ Apply change that leads to best improvement in score
    ‣ Reiterate
  ‣ Stop when no modification improves score
‣ *Pitfalls*:
  ‣ Local Maxima
  ‣ Plateaus
‣ *Solution*:
  ‣ Tabu
  ‣ Random restart
  ‣ Simulated annealing

Random start

Score1

Random change

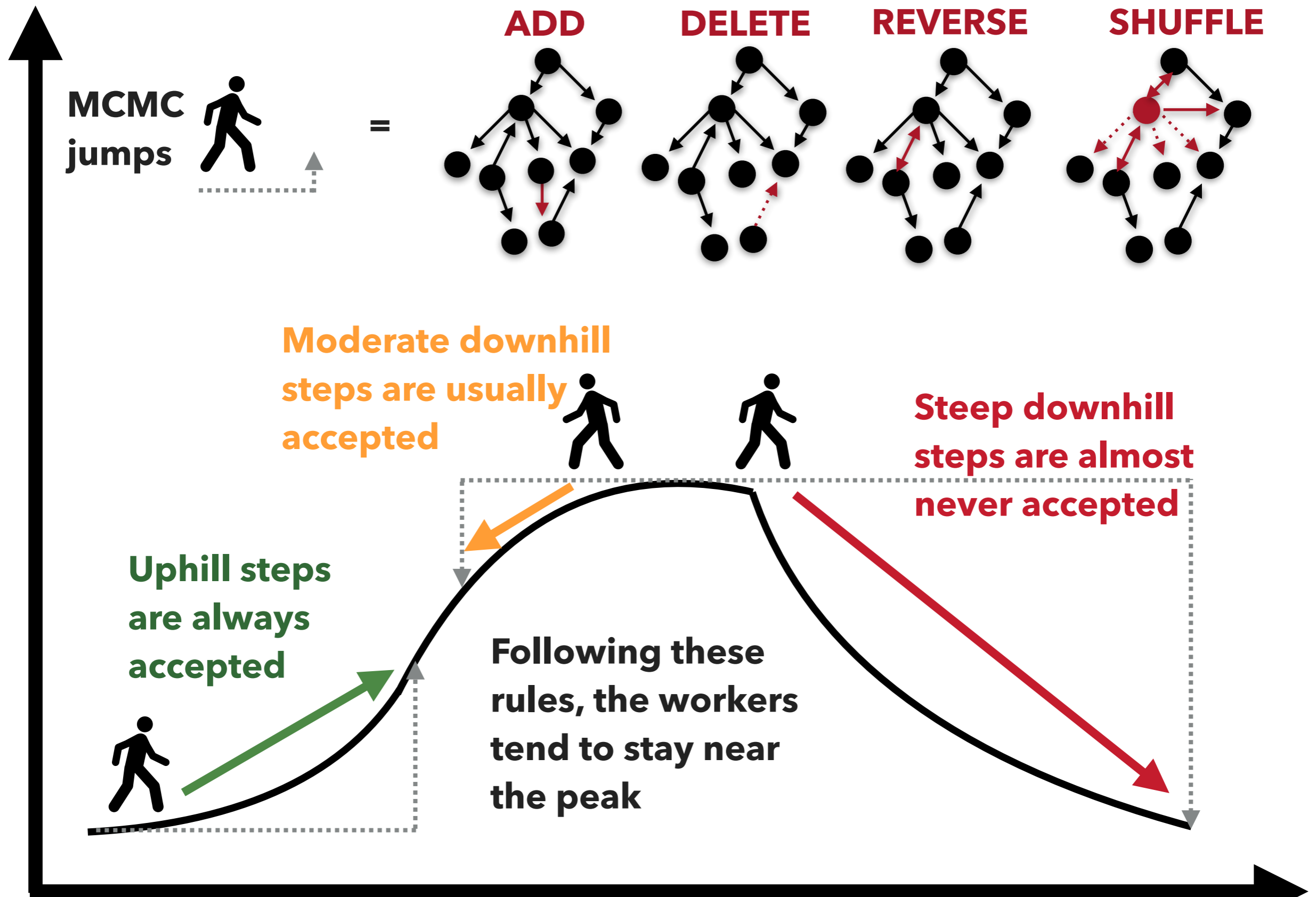Score2>Score1
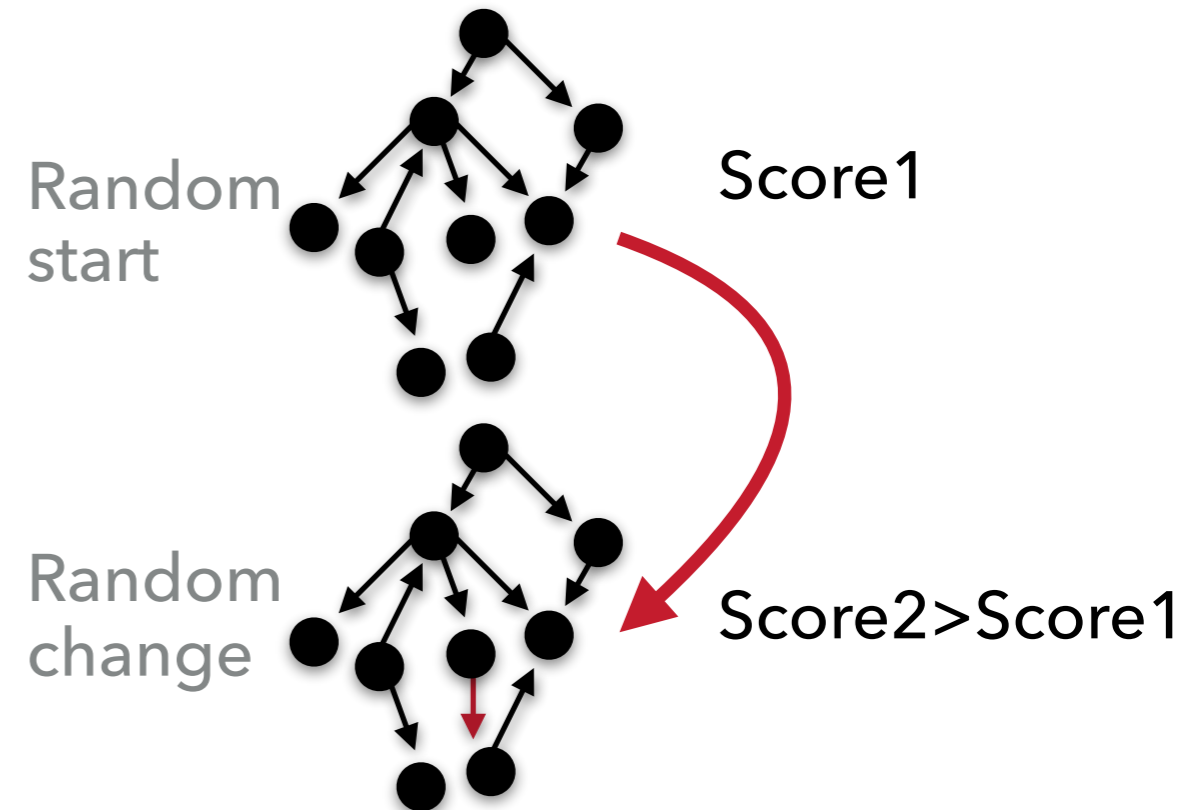
HEURISTIC SEARCH

```r
num.searches <- 200
max.steps <- 150
heur.res <- quiet(search.heuristic(score.cache = mycache,
                    score = "mlik",
                    data.dists = dist,
                    max.parents = 4,
                    start.dag = "random",
                    num.searches = num.searches,
                    max.steps = max.steps,
                    seed = 3213,
                    verbose = TRUE,
                    algo = "hc"))
```
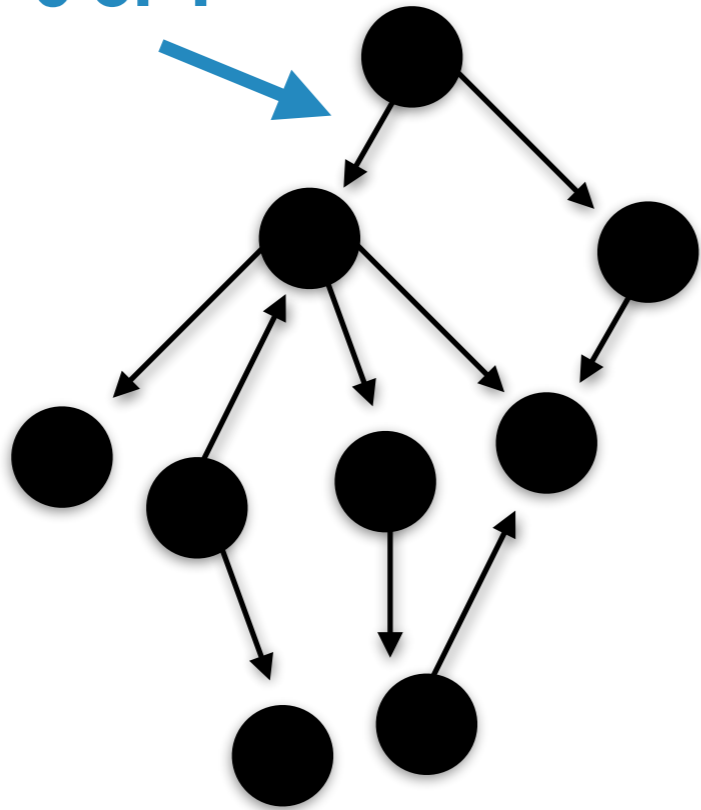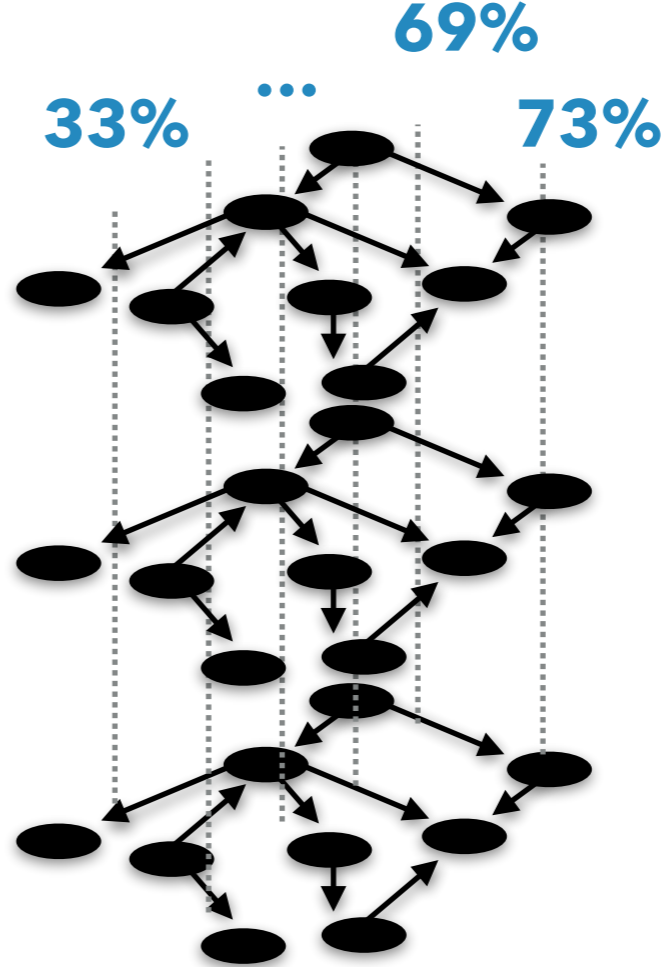
MCMC OVER STRUCTURES

ADD   **DELETE**   **REVERSE**   **SHUFFLE**

MCMC jumps =

Moderate downhill steps are usually accepted

Steep downhill steps are almost never accepted

Uphill steps are always accepted

Following these rules, the workers tend to stay near the peak

Adapted Lewis (2011)

Best Unique Bayesian Network

0 or 1

Counting prevalence of each ARC

33% ... **69%** **73%**

MCMC over structures

‣ Selecting the most probable structure

‣ Controlling for overfitting

‣ Sampling the landscape of high scoring structures

   ‣ In applied perspective avoid reducing the richness of BN modelling to only **one** structure

   ‣ Quantify the marginal impact of relationships by marginalising out over structures

HANDS-ON EXERCICES