

Can Age Explain Macroevolution?

Murat Tuğrul¹, Stephanie Keller-Schmidt², Víctor M. Eguíluz¹,
Emilio Hernández-García¹ and Konstantin Klemm²

¹IFISC (Inst. for Cross-Disciplinary Physics and Complex Syst.), UIB-CSIC, Palma de Mallorca, Spain,
²Bioinformatics Group, Dept. of Computer Sci., University of Leipzig, Germany



MOTIVATION: FINDING A BRANCHING MODEL CONSISTENT WITH AVAILABLE PHYLOGENETIC DATA

- Thinking Macroevolution (speciation) as tree-like
- Molecular Phylogeny → Phylogenetic trees (databases like TreeBASE, PANDIT, etc.)

I think

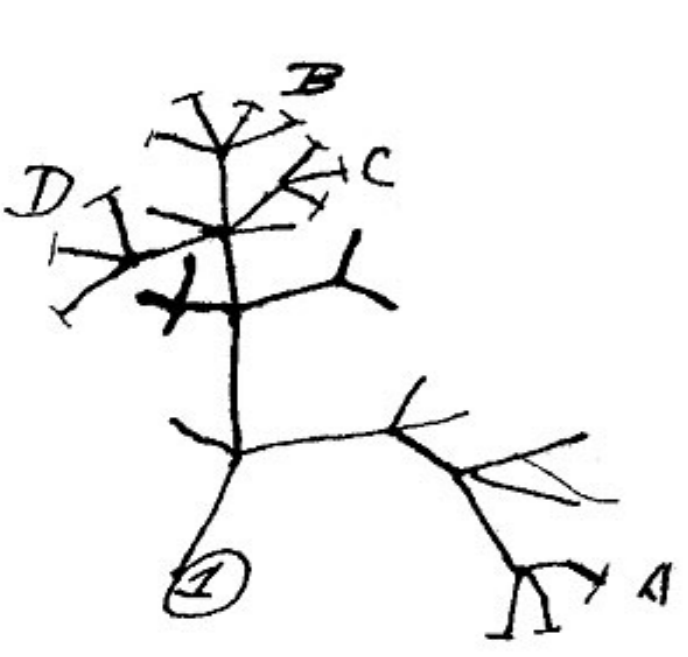


Figure: Darwin's sketch (1837)

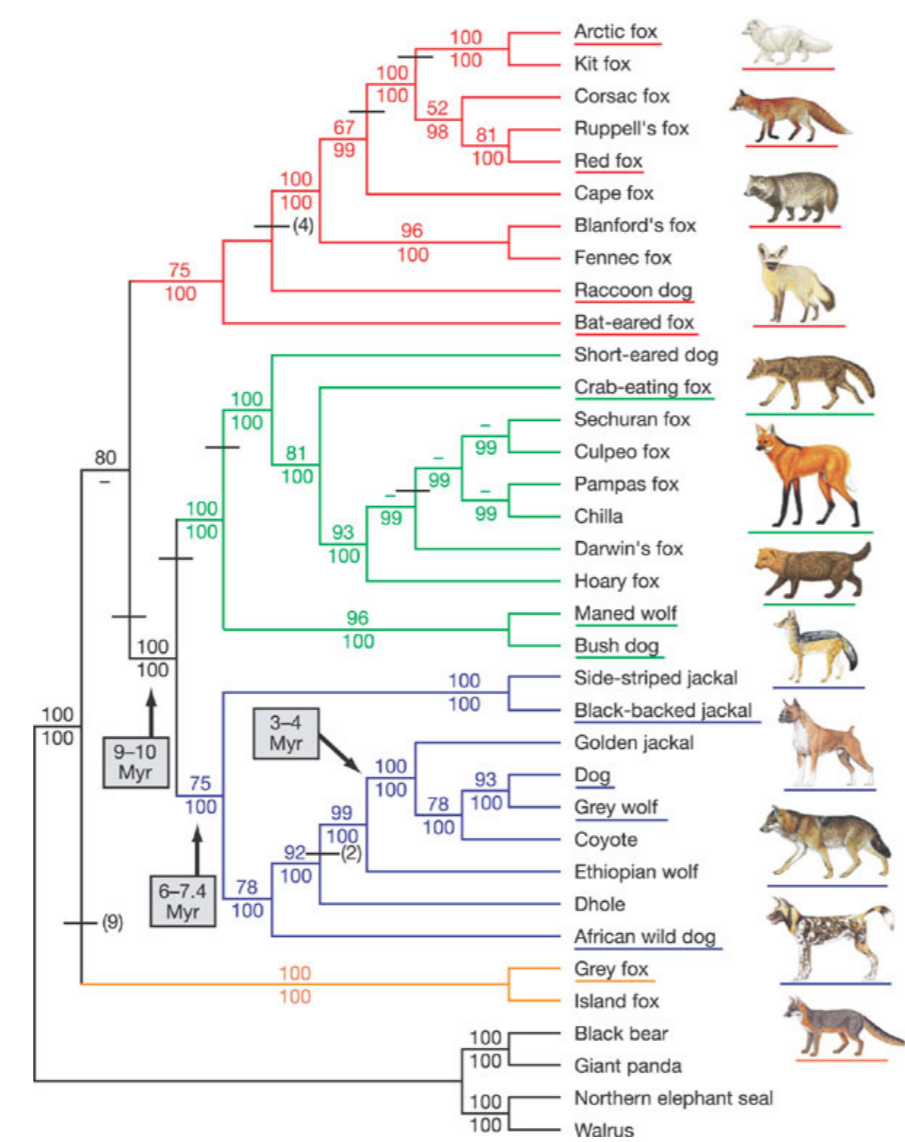


Figure: A phylogenetic tree represents the evolutionary relationship between Domestic Dogs (Pollinger et al., 2005)

- Analysing the structures of estimated phylogenetic trees and comparing to those obtained from branching models is an interesting approach to capture the rules of macroevolution.

PHYLOGENETIC TREE ANALYSIS.

- Balance notion can be used for shape analyses

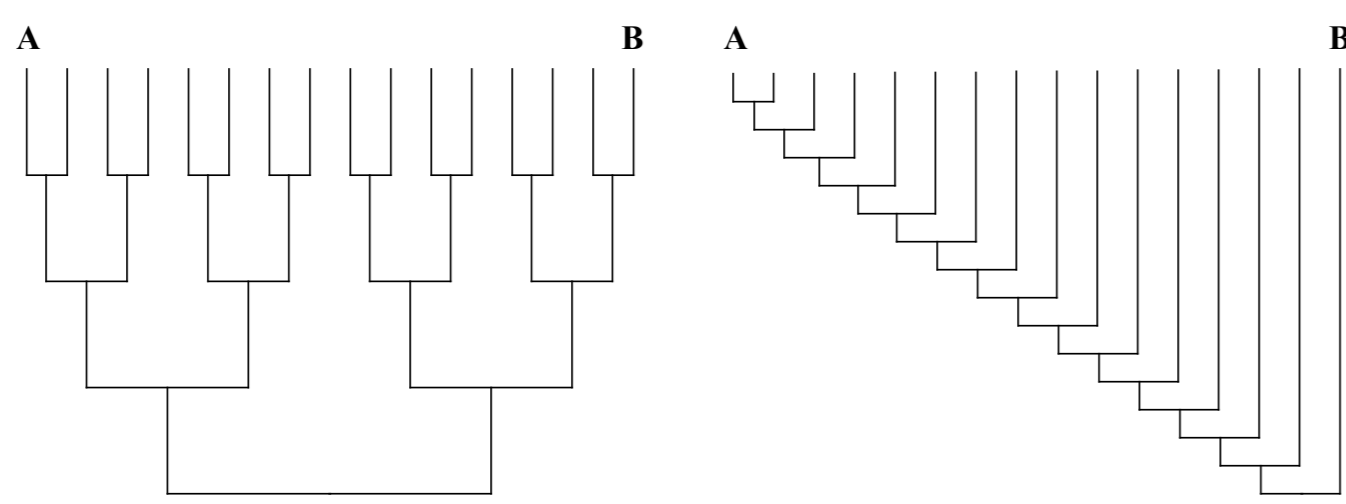


Figure: Most balanced and unbalanced binary tree of size 16 leaves.

- We use the *mean depth* of the tree, i.e. the *average number of ancestor nodes from tips to the root*:

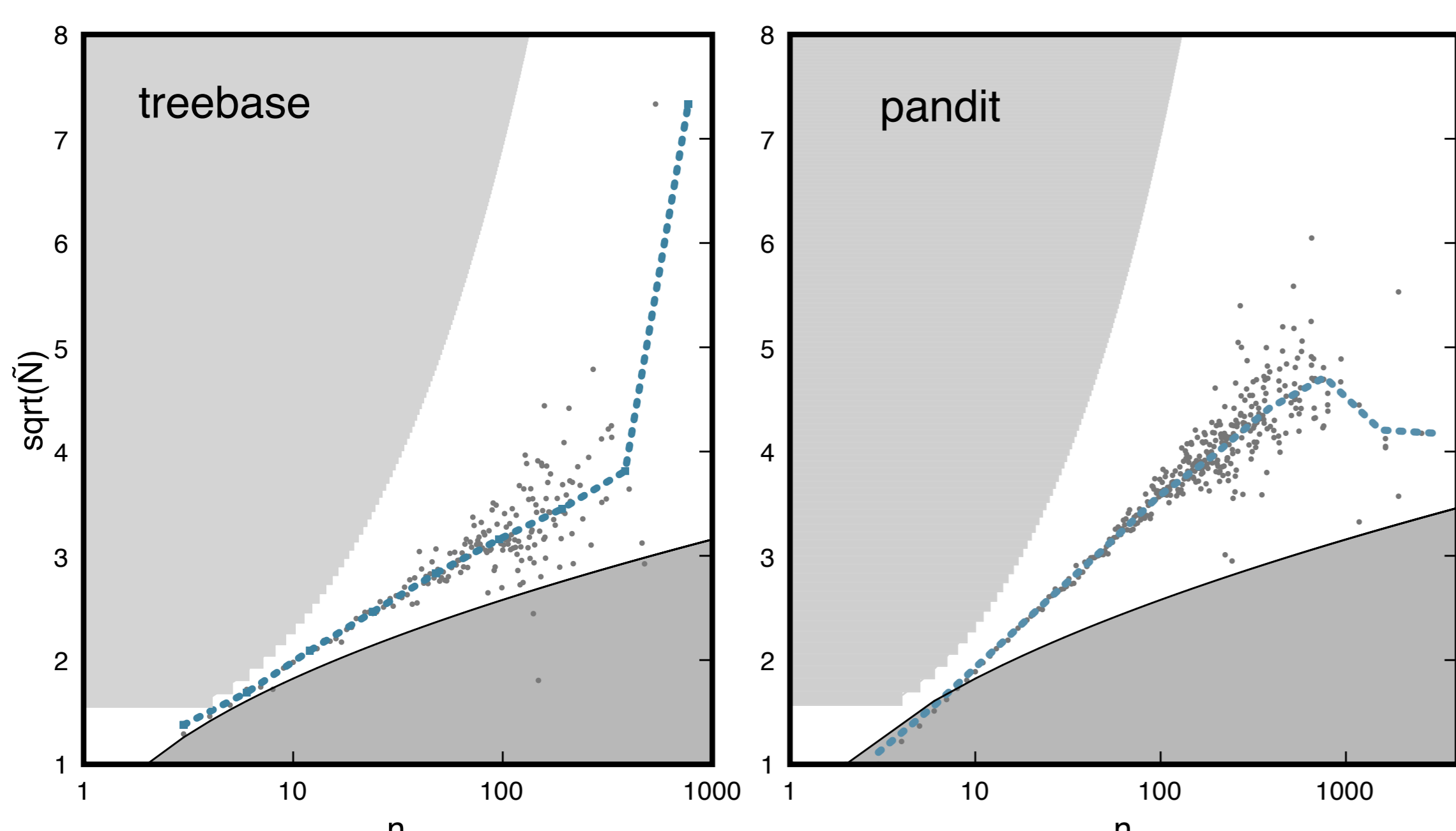
$$\bar{N} = \left(\sum_{i=1}^n N_i \right) / n$$

where N_i is the depth (the number of ancestors) of tip i , and n is the total number of tips in tree.

- We focus on the scaling behavior of this mean depth for comparison purposes. It is logarithmic for most balanced and ERM (the Equal Rates Markov) trees while linear for totally unbalanced trees.

DATABASE TREES SHOW NON-LOGARITHMIC SCALING (NOT CAPTURED YET WITH A BIOLOGICALLY MEANINGFUL BRANCHING MODEL)

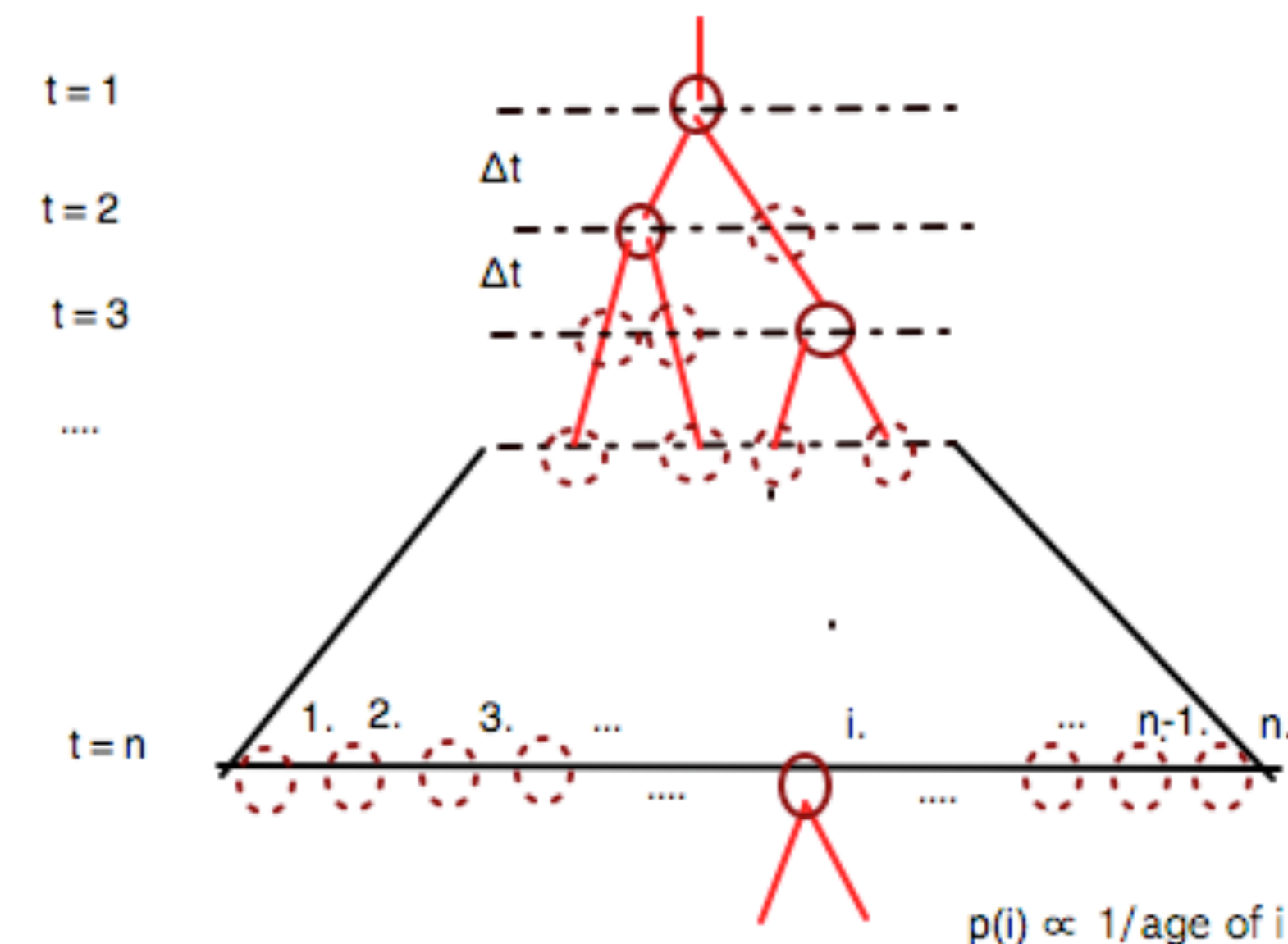
- The mean depth vs. size of phylogenetic trees contained in databases of species (TreeBASE) and proteins (PANDIT) (black dots).
- The white accessible area is limited by the extreme balanced and unbalanced binary tree behavior.
- In this scale, the behavior $\bar{N} \sim (\log n)^2$ would be a straight line.



WE PROPOSE AN AGE DEPENDENT BRANCHING MODEL

A discrete time ($\Delta t = 1$) branching model where choosing a tip to branch is inversely proportional to its age, i.e.

$$p_i = c (\text{age of } i)^{-1} ; c \text{ is normalization constant}$$



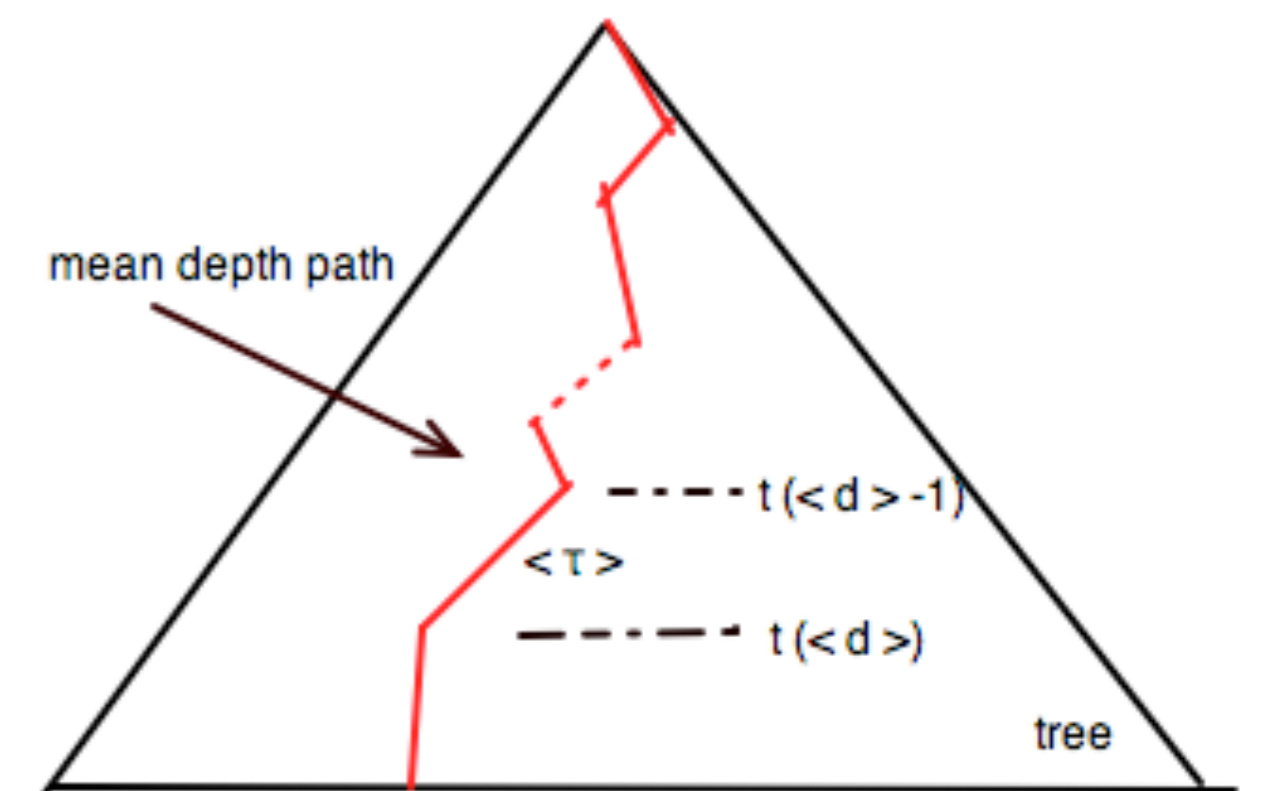
SCALING RESULT IS ANALYTICALLY OBTAINABLE FOR ASYMPTOTIC CASE

The premise: mean depth path is made of branches whose ages are approximated by expected value at time t :

$$t(\bar{N}) - t(\bar{N} - 1) = \langle \tau \rangle_{t(\bar{N})}$$

which yields to

$$\frac{d\bar{N}}{dn} = [\langle \tau \rangle_n]^{-1}.$$



Assuming a mean value for normalization constant $c(n)$, one can show

$$\langle \tau \rangle_n = n c(n)$$

Considering $c_{\text{slowest}}^{-1}(n) \leq c^{-1}(n) \leq c_{\text{fastest}}^{-1}(n)$ where c_{slowest} and c_{fastest} refer to the normalization constant of the slowest and fastest realization of the branching process:

$$c_{\text{slowest}}^{-1}(n) = 1 + \sum_{i=1}^{n-1} i^{-1}, \quad c_{\text{fastest}}^{-1}(n) = 2 \sum_{i=1}^{\lfloor n/2 \rfloor} i^{-1} + r(n)$$

where $r(n) = ((n+1)/2)^{-1}$ if n is odd, and $r(n) = 0$ otherwise,

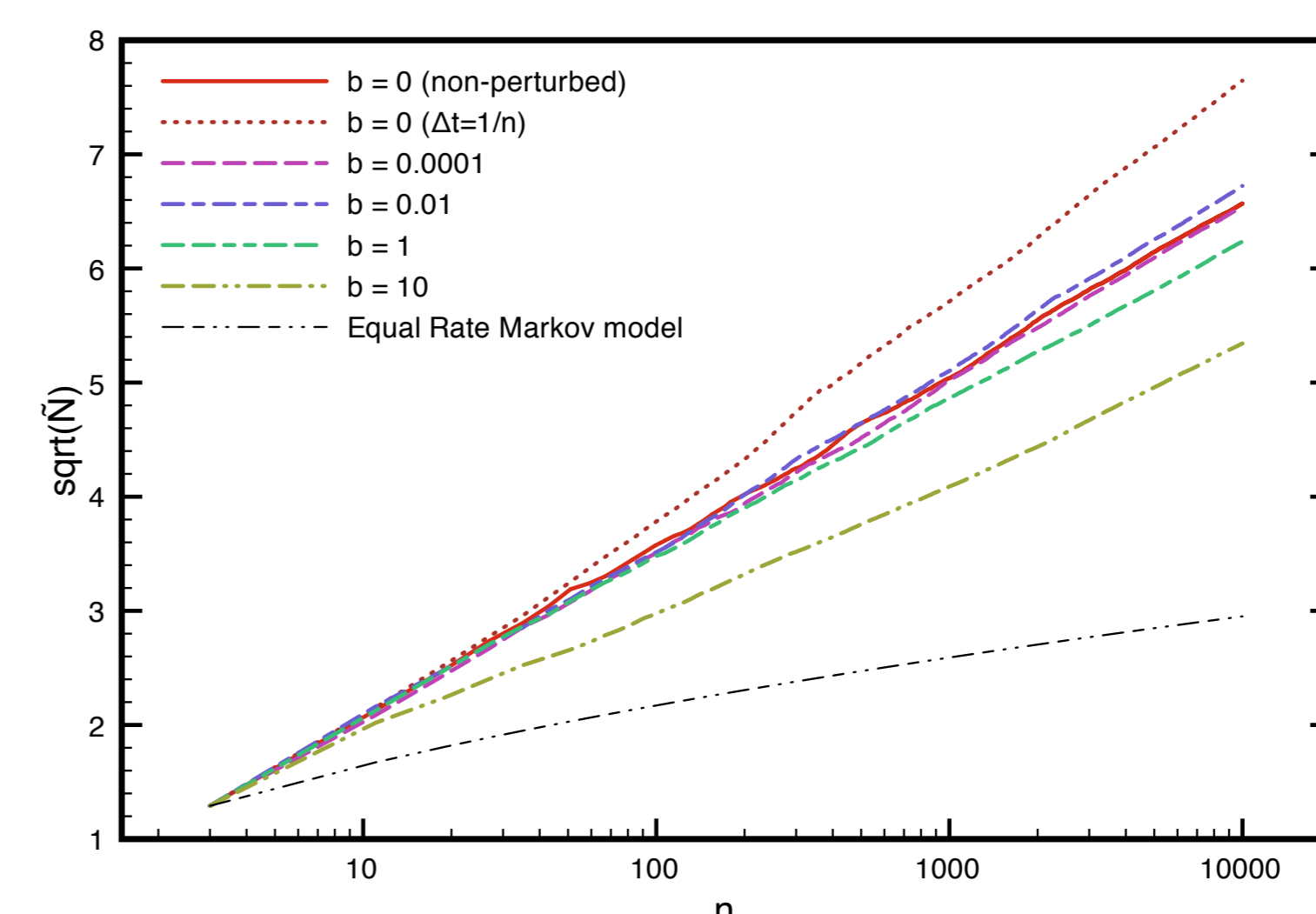
$$c(n) \rightarrow (\log n)^{-1} \text{ as } n \rightarrow \infty$$

and $\langle \tau \rangle_n \rightarrow \frac{n}{\log n}$ as $n \rightarrow \infty$.

Thus, $\bar{N} \sim (\log n)^2$ as $n \rightarrow \infty$.

WE VERIFIED THE RESULT COMPUTATIONALLY AND INVESTIGATED THE ROBUSTNESS OF MODELING

- Simulation results verify the $(\log n)^2$ behavior.
- The modeling is robust under modifications
 - from $p_i \propto (\text{age})^{-1}$ to $\propto (\text{age} + \text{const.})^{-1}$
 - from $\Delta t = 1$ to $\Delta t = 1/n$.



CONCLUSION:

We have proposed an Age-Dependent branching model with depth scaling behavior consistent with the one in estimated phylogenetic trees. See also poster 23 by Keller-Schmidt for further comparison.