

Supplementary results for: A meta-analysis of computational biology benchmarks reveals publication bias affects on speed and accuracy

Paul P. Gardner^{1,2*}, James Paterson^{1,2}, Fatemeh Ashari Ghomi^{1,2}, Sinan Uğur Umu^{1,2}, Stephanie McGimpsey^{1,2}

Abstract

Keywords

computational biology — accuracy — benchmarks — meta-analysis — software development

¹School of Biological Sciences, University of Canterbury, Christchurch, New Zealand.

²Biomolecular Interaction Centre and the Bio-Protection Research Centre, University of Canterbury, Christchurch, New Zealand.

*Corresponding author: paul.gardner@canterbury.ac.nz

Introduction

XXXX

R code for Figure 1A&B

```
#read data
d <- read.table("meanRankSpeedData.tsv", header=T)

#initialize matrices
dNames <- c("JIF", "H5", "relAge", "yearPublished", "cites", "relCites", "mindex", "hindex", "accuracyRank", "speedRank")
pNames <- c("JIF", "H5", "Rel. age", "Year", "Cites", "Rel. cites", "M-index", "H-index", "Accuracy", "Speed")
pvalMatrix<-matrix(1, length(dNames), length(dNames))
rhoMatrix <-matrix(0, length(dNames), length(dNames))
sigMatrix <-matrix("",length(dNames), length(dNames))

colnames(pvalMatrix)<-pNames
rownames(pvalMatrix)<-pNames
colnames(rhoMatrix) <-pNames
rownames(rhoMatrix) <-pNames
colnames(sigMatrix) <-pNames
rownames(sigMatrix) <-pNames

#loop through pairwise combinations, record rho and P-values
for(i in 1:length(dNames)){
  for(j in 1:length(dNames)){
    spear<-cor.test(d[,dNames[i]] == colnames(d), d[,dNames[j]] == colnames(d), method = "spearman", exact = T)  ##, alternative = "less")
    pvalMatrix[i,j] <- spear$p.value
    rhoMatrix[i,j] <- spear$estimate
    if(spear$p.value < 0.05){
      sigMatrix[i,j] <- "X"
    }
  }
}

#generate plots
pdf(file= "../figures/spearmanHeatmap.pdf", width = 7, height = 6)
par(mar = c(8,4,4,4) + .1) #c(bottom, left, top, right). default: c(5, 4, 2) + 0.1
heatmap.2(rhoMatrix, cellnote=sigMatrix,notece=1.5,notecol="black", col=redblue(40), density.info="none", trace="none", dendrogram=c("column"), symm=F,symkey=T,symbreaks=T,
scale="none", key.title = "", srtRow=45, adjRow=c(0, 1), srtCol=45, adjCol=c(1,1), breaks=(-20:20)/20,
margins = c(8, 8), cexRow=1.5, cexCol=1.5)
dev.off()

relCitesA<-cor.test(1-d$accuracyRank, as.numeric(d$relCites), method = "spearman")
hindexA <-cor.test(1-d$accuracyRank, as.numeric(d$hindex), method = "spearman")
mindexA <-cor.test(1-d$accuracyRank, as.numeric(d$mindex), method = "spearman")
H5A <-cor.test(1-d$accuracyRank, as.numeric(d$H5), method = "spearman")
relAgeA <-cor.test(1-d$accuracyRank, as.numeric(d$relAge), method = "spearman")
speedA <-cor.test(1-d$accuracyRank, as.numeric(d$speedRank), method = "spearman")
citesA <-cor.test(1-d$accuracyRank, as.numeric(d$cites), method = "spearman")
JFA <-cor.test(1-d$accuracyRank, as.numeric(d$JIF), method = "spearman")
yearA <-cor.test(1-d$accuracyRank, as.numeric(d$yearPublished),method = "spearman")

pdf(file= "../figures/spearmanBarplot.pdf", width = 5, height = 3)
op<-par(mfrow=c(1,1),cex=1.0,las=2)
barplot(t(c(mindx$estimate, hindex$estimate, relAge$estimate, H5$estimate, speed$estimate, cites$estimate, relCites$estimate, year$estimate, JIF$estimate)), names=c("M-index", "H-index", "Rel. age", "JH5", "Speed", "Cites", "Rel. cites", "Year", "JIF"), ylab="Spearman's rho",ylim=c(-0.1,0.1),main="Correlates with accuracy rank")
lines(c(-100,100),c(0,0))
dev.off()
```

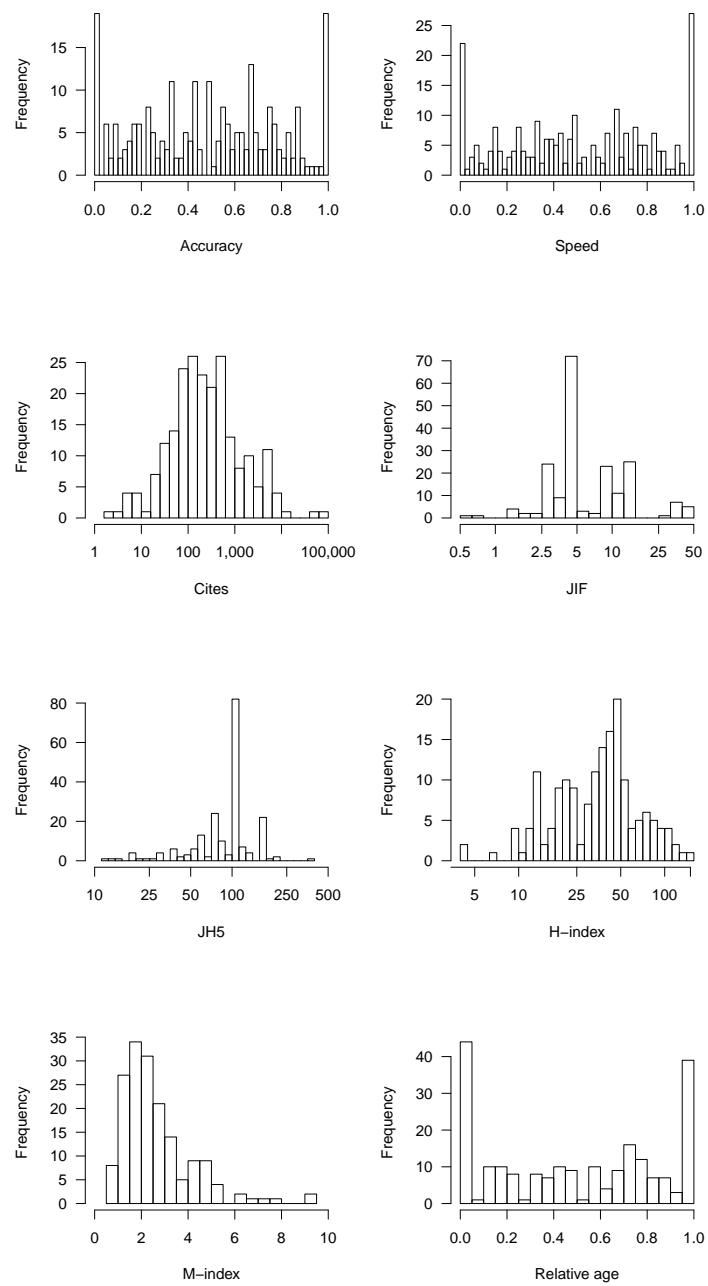


Figure 1. The distributions for the metrics used in this study. These are, reading from left to right, top to bottom: Accuracy – the mean normalised accuracy rank for each benchmarked method; Speed – the mean normalised speed rank for each benchmarked method; Cites – the number of citations to the most cited manuscript describing a method, data from GoogleScholar; JIF – the Journal Impact Factor to the highest impact journal that has published a manuscript describing a method, data from 2014 Thompson-Reuters Citation Reports; JH5 – the Journal H5 index to the highest impact journal that has published a manuscript describing a method, data from GoogleScholar 2015 Metrics; H-index – the H-index for the highest profile corresponding author from the manuscripts describing a method, data from GoogleScholar User Profiles; M-index – the M-index ($H\text{-index}/(\# \text{years since first publication})$) for the highest profile corresponding author from the manuscripts describing a method, data from GoogleScholar User Profiles;

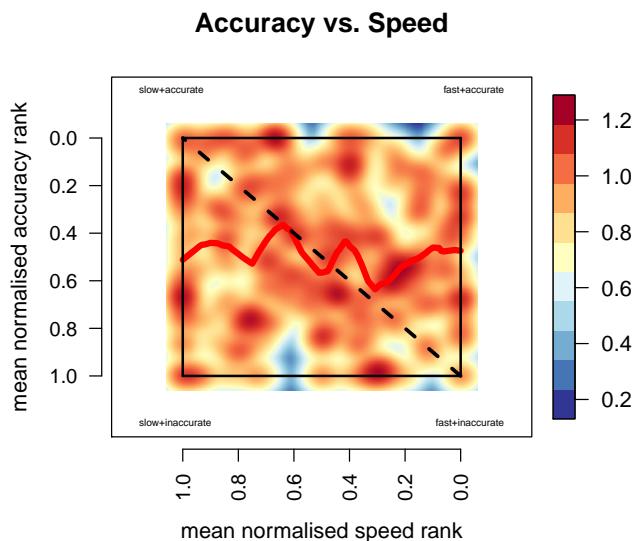


Figure 2. ...

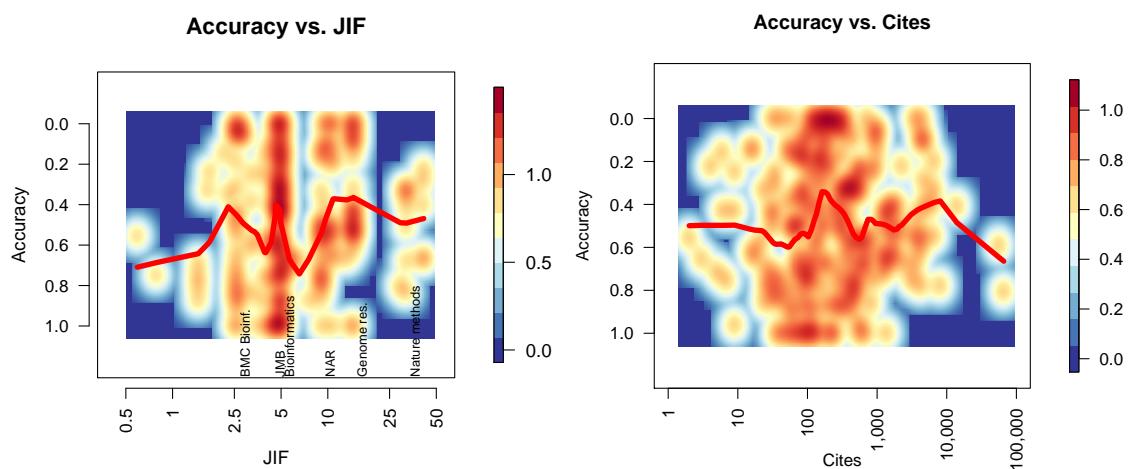


Figure 3. ...

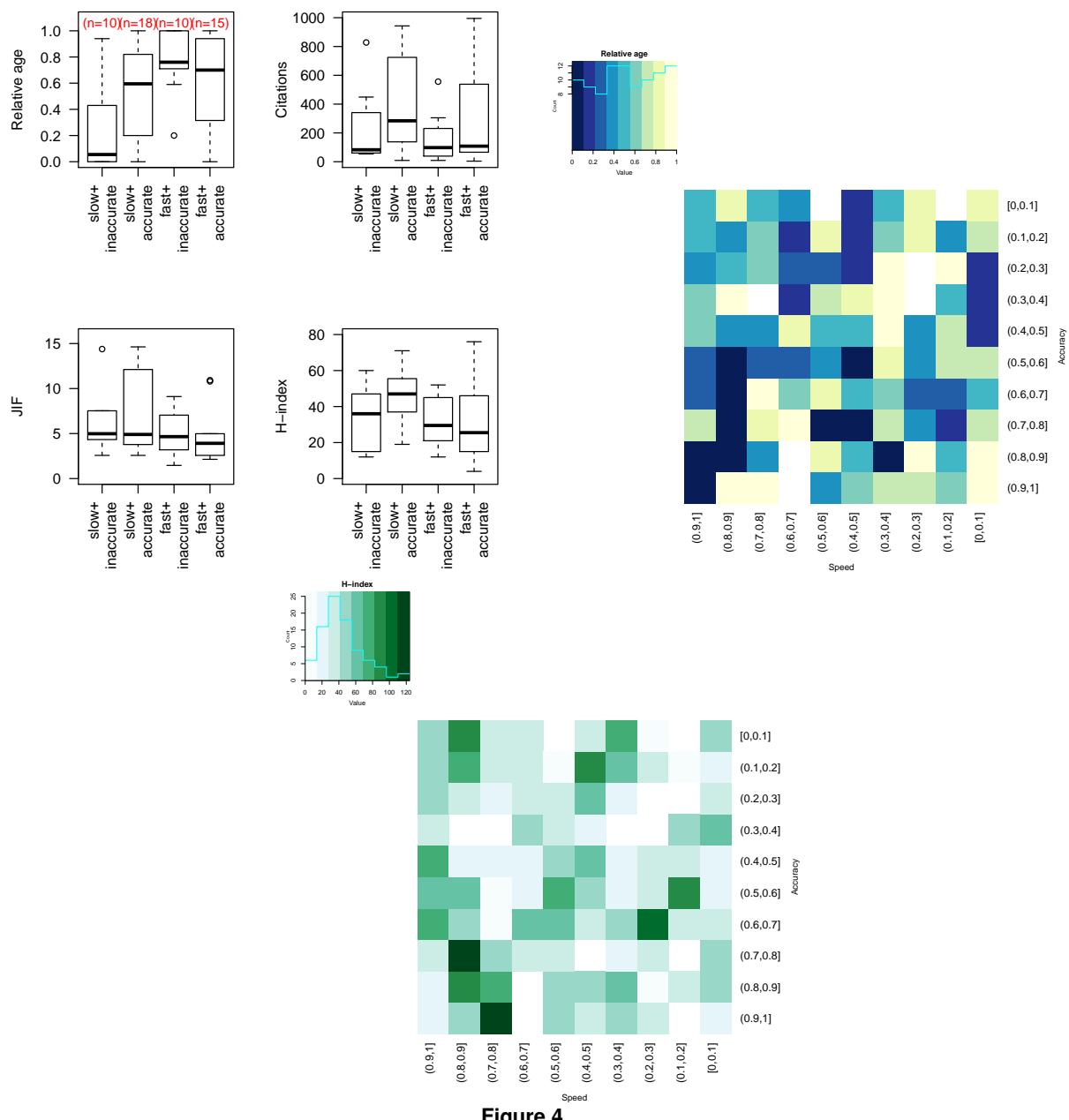


Figure 4. ...

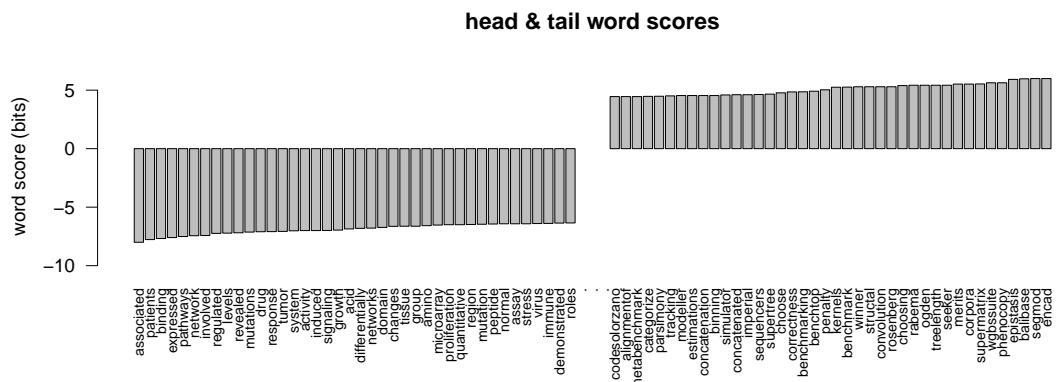


Figure 5. ...