Learning kernels from biological networks by maximizing entropy (supplementary information)

Koji Tsuda[*+] and William Stafford Noble[†]

[*]Max Planck Institute for Biological Cybernetics
Tübingen, Germany
[+]AIST Computational Biology Research Center
Tokyo Japan
[†]Department of Genome Sciences
University of Washington
Seattle, WA, USA
`koji.tsuda@tuebingen.mpg.de, noble@gs.washington.edu`

March 12, 2004

1 Functional categories from the CYGD

Tables 1 and 2 list the CYGD functional categories used in the experiments. The number of proteins in that category in the metabolic network ("M") and the protein interaction network ("I") is also shown.

2 ROC scores for other settings of C

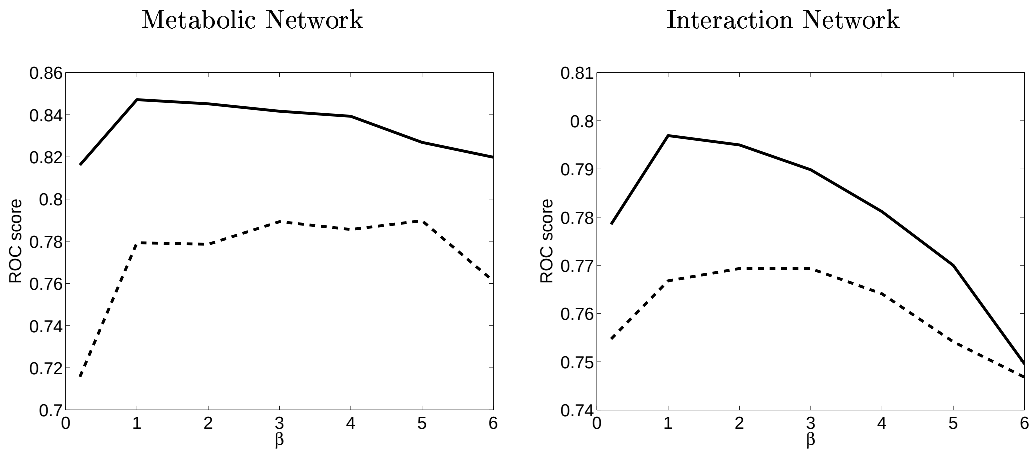
The ROC scores from the experiments with $C = 100, 10000, 100000$ are described in Figures 1,2, and 3, respectively.

Table 1: **Functional categories from the CYGD.** The table lists the CYGD functional categories used in the experiments. The number of proteins in that category in the metabolic network ("M") and the protein interaction network ("I") is also shown.

Functional category	M	I
METABOLISM	479	590
amino acid metabolism	115	131
amino acid biosynthesis	89	99
nitrogen and sulfur metabolism	—	38
nucleotide metabolism	70	102
purine ribonucleotide metabolism	34	38
C-compound and carbohydrate metabolism	182	220
C-compound and carbohydrate utilization	176	149
regulation of C-compound and carbohydrate utilization	—	72
lipid, fatty-acid and isoprenoid metabolism	75	90
lipid, fatty-acid and isoprenoid biosynthesis	49	60
metabolism of vitamins, cofactors, and prosthetic groups	48	59
biosynthesis of vitamins, cofactors, and prosthetic groups	41	50
ENERGY	133	136
respiration	—	37
CELL CYCLE AND DNA PROCESSING	78	393
DNA processing	—	161
DNA synthesis and replication	—	62
DNA recombination and DNA repair	—	95
DNA repair	—	58
cell cycle	58	290
mitotic cell cycle and cell cycle control	46	238
meiosis	—	55
TRANSCRIPTION	60	436
rRNA transcription	—	73
rRNA processing	—	46
tRNA transcription	—	56
mRNA transcription	—	304
mRNA synthesis	—	208
general transcription activities	—	53
transcriptional control	—	155
mRNA processing splicing, 5'-, 3'-end processing	—	93
splicing	—	65
PROTEIN SYNTHESIS	48	265
ribosome biogenesis	—	165
translation	—	47
aminoacyl-tRNA-synthetases	36	35

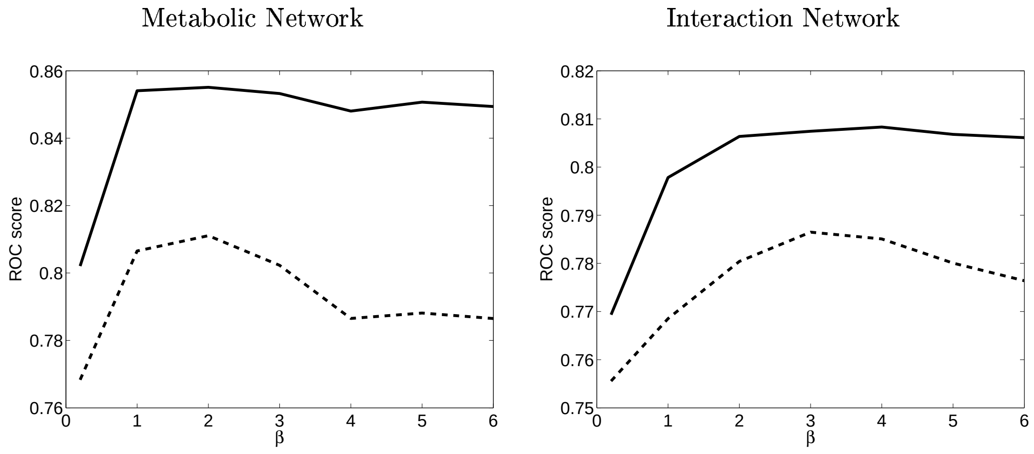
Table 2: **Functional categories from the CYGD (continued).** See caption for Table 1.

Functional category	M	I
PROTEIN FATE folding, modification, destination	70	316
protein targeting, sorting and translocation	—	74
protein modification	52	95
assembly of protein complexes	—	60
proteolytic degradation	—	89
cytoplasmic and nuclear degradation	—	62
CELLULAR TRANSPORT AND TRANSPORT MECHANISMS	40	253
nuclear transport	—	45
vesicular transport — Golgi network, etc	—	77
vacuolar transport	—	33
cellular import	—	42
CELL RESCUE, DEFENSE AND VIRULENCE	40	114
stress response	—	85
REGULATION OF / INTERACTION WITH CELLULAR ENVIRONMENT	43	88
ionic homeostasis	30	55
homeostasis of cations	30	47
cellular sensing and response	—	34
chemoperception and response	—	34
CELL FATE	46	234
cell growth / morphogenesis	—	53
cell differentiation	42	214
fungal cell differentiation	42	214
budding, cell polarity and filament formation	—	108
pheromone response, mating-type determination, sex-specific proteins	—	92
sporulation and germination	—	62
CONTROL OF CELLULAR ORGANIZATION	—	100
plasma membrane	—	51
cytoplasm	177	419
cytoskeleton	—	79
endoplasmic reticulum	31	90
Golgi	—	41
nucleus	75	501
mitochondrion	99	192
vacuole or lysosome	—	36
SUBCELLULAR LOCALISATION	441	1369
TRANSPORT FACILITATION	38	78
ion transporters	—	40
cation transporters	—	32
transport mechanism	35	—
transport ATPases	35	—
CLASSIFICATION NOT YET CLEAR-CUT	—	45


Metabolic Network

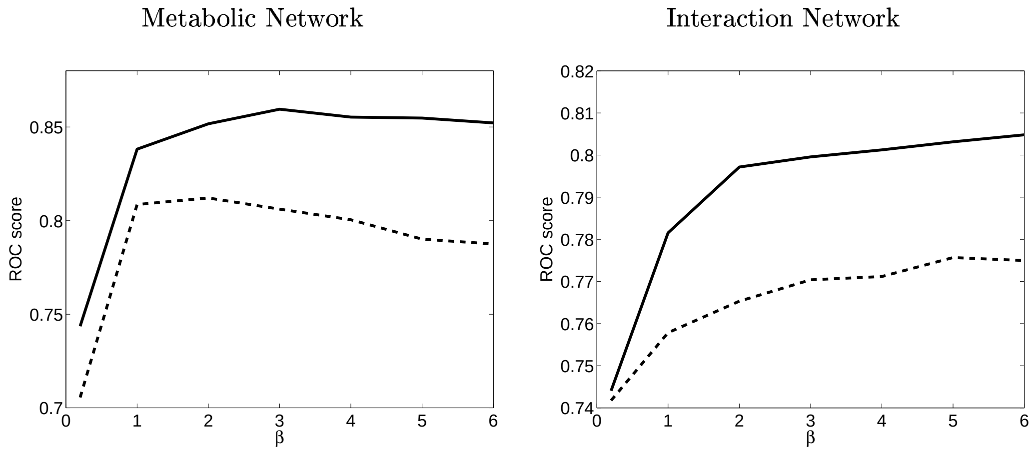
Interaction Network

Figure 1: $C = 100$


Metabolic Network

Interaction Network

Figure 2: $C = 10000$


Metabolic Network

Interaction Network

Figure 3: $C = 100000$