Link Enrichment for Diffusion-based Graph Node Kernels

Dinh Tran-Van[1], Alessandro Sperduti[1], and Fabrizio Costa[2]

[1] Department of Mathematics, Padova University
[2] Department of Computer Science, University of Exeter
{dinh, sperduti}@math.unipd.it, f.costa@exeter.ac.uk

APPENDIX

In this document, we present in detail the results obtained from the experiment described on the paper. We report the performance on 14 disease gene classes and four different biological graphs induced by BioGPS, Biogridphys, HPRD and Omim. For each disease class and graph, we show the average AUC(%) for each diffusion-based graph node kernel in two cases: plain diffusion kernel (denoted by "-" and followed by kernel natation) and diffusion kernel on a modified graph obtained by link enrichment process (denoted by + followed by nation of kernel which is used for link enrichment).

Table 1. Indices of genetic disease classes

Index	Disease
1	Cancer
2	Cardiovascular
3	Connective
4	Dermatological
5	Developmental
6	Endocrine
7	Hematological
8	Immunological
9	Metabolic
10	Muscular
11	Neurological
12	Ophthalmological
13	Renal
14	Skeletal

Table 2. Notations of graph node kernels

Notaion	Kernel
A	LEDK
B	MEDK
C	MDK
D	RLK
E	CDNK

Table 3. *Predictive performance on 14 gene-disease associations using network induced by the BIOGPS*

Disease	-A	+A	+B	+C	+D	+E	-B	+A	+B	+C	+D	+E
1	61.8	63.8	62.5	63.7	63.9	63.4	57.9	62.8	62.8	62.2	62.6	62.2
2	55.6	67.5	62.3	70.4	69.5	63.5	52.0	62.8	57.3	57.9	60.1	58.3
3	50.3	62.3	61.2	63.6	61.9	60.0	49.5	57.0	55.1	56.1	57.2	55.2
4	61.3	73.9	73.1	73.3	73.7	73.9	62.7	72.7	67.9	72.5	74.7	68.5
5	55.7	62.8	60.6	63.2	62.9	62.9	55.1	61.7	61.0	62.9	62.6	60.5
6	60.1	68.8	66.8	70.4	69.4	67.4	59.7	66.1	62.2	68.2	67.0	63.4
7	68.8	73.6	73.0	73.6	74.2	73.4	67.7	73.5	73.1	74.2	74.8	72.1
8	71.1	77.9	72.3	76.5	77.8	73.2	65.6	71.7	70.8	70.4	71.8	70.9
9	62.4	63.8	63.1	64.4	64.3	63.8	63.7	66.9	65.7	67.2	66.7	66.5
10	69.4	72.9	71.2	74.5	72.6	71.8	62.5	73.6	70.4	72.7	71.9	71.1
11	60.0	63.9	63.1	64.6	63.6	63.3	58.4	61.4	61.9	61.4	61.8	61.7
12	64.4	76.2	73.5	76.1	75.2	73.2	61.9	71.2	70.3	71.7	72.3	70.5
13	57.2	66.1	62.4	65.8	63.0	59.3	55.1	63.7	63.5	62.4	61.8	60.6
14	55.4	61.7	61.7	61.6	63.0	60.1	55.0	63.7	64.8	62.1	63.1	62.5
$\overline{AUC}$	61.0	68.2	66.2	68.7	68.2	66.4	59.1	66.3	64.8	65.8	66.3	64.6

Table 4. *Predictive performance on 14 gene-disease associations using network induced by the BIOGPS.*

Disease	-C	+A	+B	+C	+D	+E	-D	+A	+B	+C	+D	+E
1	60.4	65.2	63.6	64.7	65.3	64.6	61.1	63.0	62.2	63.0	63.3	62.5
2	53.0	67.5	64.2	66.8	66.9	65.1	54.4	63.7	59.7	59.8	63.8	61.0
3	50.4	62.7	52.8	61.8	62.3	57.1	50.4	57.4	56.7	56.9	58.6	56.1
4	60.2	74.0	72.4	72.8	74.3	74.9	61.9	72.0	68.2	69.6	71.7	69.9
5	54.8	60.8	61.1	61.4	61.5	61.9	54.8	61.2	60.2	61.9	61.5	61.3
6	62.0	70.6	67.6	70.8	69.5	66.1	61.3	70.0	67.0	69.3	69.5	67.6
7	66.0	73.4	71.7	72.5	72.9	73.2	69.8	73.7	73.3	73.7	74.1	73.4
8	71.4	76.9	73.9	75.3	76.0	74.0	68.7	75.0	73.3	74.4	74.2	72.8
9	59.3	64.2	62.1	63.6	63.3	62.4	62.4	64.4	63.7	65.0	64.7	64.3
10	69.6	75.9	74.4	76.1	76.2	73.9	68.4	72.8	70.2	73.5	71.8	70.5
11	56.0	61.4	58.1	61.7	61.9	60.0	60.6	63.5	63.1	63.5	63.6	62.8
12	65.4	74.2	72.7	73.6	74.2	74.0	64.5	76.2	73.4	74.3	74.6	74.3
13	57.3	67.2	60.5	67.3	67.7	63.5	56.4	63.3	61.7	61.7	63.3	60.5
14	55.5	63.1	60.0	62.3	63.8	63.2	54.9	62.0	61.8	62.0	62.4	61.1
$\overline{AUC}$	60.1	68.4	65.4	67.9	68.3	66.7	60.7	67.0	65.3	66.3	66.9	65.6

Table 5. *Predictive performance on 14 gene-disease associations using network induced by the BIOGRIDphys.*

Disease	-A	+A	+B	+C	+D	+E	-B	+A	+B	+C	+D	+E
1	76.4	79.2	77.8	77.8	79.6	78.3	73.2	79.7	75.0	73.9	79.5	74.9
2	60.5	66.4	61.6	60.8	67.3	63.4	52.3	63.3	55.5	53.0	61.3	53.8
3	62.2	77.2	67.7	67.4	77.2	69.2	49.1	65.2	53.7	51.8	70.0	54.2
4	69.2	79.7	75.4	75.0	79.1	76.3	63.5	77.5	73.9	71.9	77.9	72.7
5	59.2	66.6	63.3	62.1	67.5	62.7	50.2	59.5	56.2	53.4	58.7	55.0
6	63.1	69.7	66.7	66.7	70.1	67.1	54.4	62.9	60.3	58.7	63.0	59.0
7	60.1	67.2	62.3	61.8	72.2	61.8	53.2	64.1	60.3	59.1	67.8	59.1
8	70.6	75.1	71.0	71.1	75.7	71.8	62.9	73.6	68.0	67.4	74.5	67.6
9	69.1	72.1	70.5	70.5	72.3	73.7	61.3	69.2	63.6	62.1	69.0	64.9
10	61.3	70.4	65.6	66.7	70.4	68.1	57.3	67.4	60.8	60.3	68.6	60.3
11	58.9	64.0	61.4	60.6	64.1	62.1	56.2	61.5	59.3	58.4	61.4	59.6
12	61.0	68.3	70.8	66.3	69.5	67.4	59.6	68.6	67.6	67.1	69.4	66.7
13	55.4	66.3	64.7	57.3	69.1	61.7	57.6	71.5	67.2	66.2	71.9	68.3
14	57.4	68.7	61.4	60.9	72.4	64.3	53.2	64.7	60.0	57.8	64.5	58.9
$\overline{AUC}$	63.2	70.8	67.2	66.1	71.9	67.7	57.4	67.8	63.0	61.5	68.4	62.5

Table 6. *Predictive performance on 14 gene-disease associations using network induced by the BIOGRIDphys.*

Disease	-C	+A	+B	+C	+D	+E	-D	+A	+B	+C	+D	+E
1	66.4	75.6	69.9	72.1	79.2	73.5	76.6	80.0	78.2	78.1	80.2	78.7
2	54.4	63.6	57.5	57.8	65.9	58.7	59.3	66.5	61.8	59.5	65.9	62.8
3	60.0	74.0	60.6	60.3	74.0	65.5	64.4	76.3	68.7	68.4	76.3	71.4
4	60.2	73.3	65.1	67.3	71.8	67.3	70.1	79.5	76.9	75.8	79.7	76.7
5	48.6	61.9	56.5	55.4	62.7	57.4	58.7	66.0	62.6	60.9	64.6	61.1
6	61.7	68.8	65.2	64.5	68.8	65.2	63.2	69.5	67.1	67.1	70.1	67.3
7	56.4	65.5	58.5	58.8	68.7	60.3	61.3	66.8	63.1	62.3	71.1	62.5
8	66.8	74.0	73.1	74.0	74.6	72.5	72.3	75.3	72.3	72.3	75.4	72.7
9	74.0	78.5	75.1	75.6	79.1	75.1	70.4	73.0	71.6	71.7	72.6	74.8
10	54.4	66.9	64.1	61.3	68.2	58.7	62.3	71.0	66.5	67.3	70.8	68.3
11	58.4	64.5	60.0	60.5	63.9	61.4	59.1	62.9	61.0	60.3	63.2	61.8
12	55.9	66.9	56.9	65.4	66.7	64.4	60.8	67.2	70.1	66.1	68.5	67.1
13	54.8	66.2	64.4	58.7	69.4	63.5	55.2	65.5	64.1	57.0	68.1	60.9
14	55.1	66.6	59.7	57.5	67.6	62.6	56.4	67.5	61.7	60.7	68.7	64.1
$\overline{AUC}$	59.1	69.0	63.3	63.5	70.0	64.7	63.6	70.5	67.6	66.3	71.1	67.9

Table 7. *Predictive performance on 14 gene-disease associations using network induced by the HPRD.*

Disease	-A	+A	+B	+C	+D	+E	-B	+A	+B	+C	+D	+E
1	75.5	77.7	77.8	76.7	77.9	76.1	75.1	77.2	77.0	76.5	76.0	75.9
2	56.6	61.1	58.6	59.6	63.5	58.6	55.9	59.5	56.6	58.0	60.2	58.0
3	61.2	73.0	69.0	71.3	76.1	67.9	56.1	70.0	61.1	67.7	71.1	63.3
4	67.3	74.9	71.0	70.9	74.1	69.7	65.5	72.7	68.6	68.0	71.8	67.5
5	57.6	66.9	65.6	68.6	68.4	68.2	55.1	67.3	62.9	64.5	66.5	62.6
6	67.1	73.7	72.0	72.1	73.5	69.3	64.8	71.8	69.8	69.1	68.6	66.4
7	68.8	73.6	72.4	72.7	73.7	71.9	65.5	71.9	69.4	71.6	72.4	69.8
8	76.2	82.4	76.7	81.9	83.3	77.1	73.4	80.7	73.6	79.5	82.3	74.8
9	68.2	76.2	75.6	71.2	74.5	77.3	64.4	72.2	70.9	68.6	69.4	76.0
10	66.0	76.2	75.9	74.6	76.6	73.8	60.4	74.9	71.8	69.2	71.0	69.4
11	60.5	65.2	63.2	64.4	65.0	63.8	58.3	64.1	60.7	62.7	63.2	61.5
12	61.9	69.8	69.3	68.5	72.1	64.3	60.5	68.8	68.8	67.6	68.0	63.0
13	67.6	72.2	71.9	73.3	72.6	71.4	65.1	71.5	69.4	69.4	69.7	67.1
14	68.4	74.1	72.7	71.8	74.1	71.2	64.6	70.5	66.7	67.8	66.9	67.5
$\overline{AUC}$	65.9	72.6	70.8	71.3	73.2	70.0	63.2	70.9	67.7	68.6	69.8	67.3

Table 8. *Predictive performance on 14 gene-disease associations using network induced by the HPRD.*

Disease	-C	+A	+B	+C	+D	+E	-D	+A	+B	+C	+D	+E
1	75.6	78.2	78.0	78.8	78.6	77.1	75.6	78.2	78.1	77.7	78.8	77.7
2	58.0	62.5	60.6	60.7	63.0	60.7	58.0	60.4	59.2	60.0	63.5	59.2
3	65.5	73.4	70.7	73.5	76.6	69.4	64.4	72.0	70.1	73.7	74.6	70.2
4	68.1	73.9	73.2	72.0	74.3	71.3	68.3	75.1	72.1	71.2	74.0	71.5
5	59.0	68.1	67.1	67.2	68.3	66.9	59.0	68.5	66.7	69.0	68.2	67.8
6	68.3	75.5	72.2	75.3	73.8	71.9	67.1	73.4	72.3	72.1	73.6	70.7
7	71.1	73.5	72.7	72.8	74.2	73.1	70.1	73.3	72.3	72.9	73.5	72.4
8	77.6	82.3	78.8	81.4	83.3	79.2	79.3	82.7	80.0	82.9	83.2	80.3
9	71.0	77.5	77.1	77.2	77.5	81.0	69.8	76.0	75.6	72.5	75.3	78.8
10	67.1	77.0	69.9	73.8	77.5	74.8	69.6	76.3	75.2	76.8	77.1	75.2
11	60.3	65.2	63.4	67.2	65.6	64.4	61.3	65.8	63.5	65.1	65.1	64.5
12	59.2	71.7	70.9	68.6	73.0	64.2	61.7	70.9	69.7	69.7	72.0	64.3
13	65.3	73.3	71.1	73.3	73.3	71.8	67.7	74.1	72.5	74.0	73.2	71.7
14	63.4	72.4	67.1	67.3	73.7	72.1	68.7	74.3	72.5	71.8	74.5	73.9
$\overline{AUC}$	66.4	73.2	70.9	72.1	73.8	71.3	67.2	72.9	71.4	72.1	73.3	71.3

Table 9. *Predictive performance on 14 gene-disease associations using network induced by the OMIM.*

Disease	-A	+A	+B	+C	+D	+E	-B	+A	+B	+C	+D	+E
1	84.9	86.9	85.6	87.4	86.7	85.8	84.4	86.3	85.3	85.9	85.9	85.4
2	76.6	78.4	78.4	78.8	78.4	78.4	75.7	77.1	77.4	77.3	77.1	77.1
3	78.4	83.3	83.3	83.8	83.2	84.6	74.4	82.1	80.8	82.5	82.4	82.2
4	91.3	93.0	93.0	93.0	93.0	93.1	89.3	93.0	91.5	92.6	93.0	92.1
5	76.1	82.3	79.7	84.1	80.4	80.7	75.6	83.9	79.6	84.3	82.4	79.9
6	81.9	84.7	83.7	83.9	84.6	83.8	79.1	82.7	80.8	83.6	84.1	80.8
7	81.2	85.2	84.5	84.1	84.9	83.2	79.5	84.3	82.8	82.2	84.3	81.9
8	84.3	90.5	91.4	91.0	90.5	92.2	83.9	90.1	89.8	89.7	90.1	90.9
9	78.8	80.6	80.4	80.8	80.4	80.4	78.2	79.9	79.9	80.1	80.3	80.4
10	86.3	87.6	87.6	87.6	87.6	87.6	86.7	88.1	88.1	88.1	88.1	88.1
11	83.3	84.6	84.6	84.6	84.6	84.6	79.8	83.0	81.7	83.1	82.9	82.2
12	82.0	87.3	85.2	87.0	86.8	85.3	79.8	85.3	81.9	87.2	85.4	81.8
13	85.0	88.6	88.9	88.6	89.4	89.2	84.1	88.9	88.6	89.4	89.6	89.1
14	97.4	99.2	98.6	99.2	99.2	99.5	97.2	99.3	98.2	99.1	99.1	99.6
$\overline{AUC}$	83.4	86.6	86.1	86.7	86.4	86.3	82.0	86.0	84.8	86.1	86.0	85.1

Table 10. *Predictive performance on 14 gene-disease associations using network induced by the OMIM.*

Disease	-C	+A	+B	+C	+D	+E	-D	+A	+B	+C	+D	+E
1	87.3	89.7	88.7	90.2	89.5	88.4	84.7	86.8	85.5	86.5	86.5	85.7
2	71.2	72.9	72.1	72.1	72.8	72.1	76.4	78.0	78.0	78.2	78.0	78.0
3	78.9	84.2	83.3	84.2	84.0	84.2	77.5	82.7	82.6	83.2	82.9	83.6
4	88.8	90.2	90.0	90.2	90.0	90.0	91.4	92.9	92.8	92.8	92.9	93.1
5	77.7	82.2	80.6	83.5	83.1	82.5	76.3	82.3	80.1	83.8	81.3	80.9
6	76.3	82.4	81.9	82.0	83.7	82.2	82.2	84.9	83.8	83.8	84.6	84.5
7	82.7	85.4	84.1	85.4	85.8	84.1	80.6	84.6	83.9	84.3	83.9	83.3
8	89.3	91.6	91.3	91.8	91.7	92.4	84.3	90.5	90.1	90.5	90.3	93.5
9	78.5	81.5	81.5	81.5	81.5	81.5	78.6	80.4	80.4	80.8	80.4	80.4
10	85.1	88.0	88.0	88.1	88.3	88.0	86.0	87.5	87.5	87.6	87.5	87.5
11	82.0	82.9	82.8	82.7	82.8	82.6	83.0	84.4	84.4	84.4	84.4	84.4
12	84.7	87.8	86.5	88.2	87.7	85.9	81.3	86.5	83.8	86.4	86.3	85.0
13	78.2	83.1	83.1	83.1	83.8	83.1	85.0	88.5	88.5	88.5	89.0	88.5
14	97.5	98.7	98.2	98.4	98.6	98.7	97.4	99.2	98.3	99.2	99.2	99.5
$\overline{AUC}$	82.7	85.7	85.2	85.8	86.0	85.4	83.2	86.4	85.7	86.4	86.2	86.3