Provenance and Knowledge Abstraction for Reachability: A Framework for Knowledge Discovery

Delroy Cameron, Ph.D. Student
Ohio Center for Excellence in Knowledge Enabled Computing (Kno.e.sis)
Wright State University

Mentor: Olivier Bodenreider

Outline

- Background
- Reachability
  - Provenance Metadata
  - Knowledge Abstraction
- Approach
- Preliminary Results
- Knowledge Discovery
- Future Work
- Conclusion

Background

Search: PubMed ▼   RSS Save search Limits Advanced search Help

16804918[uid]   Search Clear

Display Settings: ☑ Abstract   Send to: ☑   FULL TEXT ONLINE WILEY InterScience

Genes Chromosomes Cancer. 2006 Sep;45(9):874-81.

Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis.

Killian A, Sarafan-Vasseur N, Sesboüé R, Le Pessot F, Blanchard F, Lamy A, Laurent M, Flaman JM, Frébourg T.

INSERM U614, Faculté de Médecine et de Pharmacie, 22 Boulevard Gambetta, 76183 Rouen, France.

Abstract
The most common form of genomic instability observed in colorectal cancer is chromosomal instability (CIN), whose molecular bases remain to be determined. We have previously demonstrated that inactivation in human cells of several components of the Pes1-Bop1 complex (BOP1, GRWD1, PES1, ORC6L, and RPL3), involved in ribosome biogenesis, altered chromosome segregation. To determine the contribution to colorectal tumorigenesis of somatic alterations of genes involved in ribosome biogenesis, we screened 56 primary colorectal cancers, using quantitative multiplex PCR of short fluorescent fragments, a sensitive method for the detection of gene dosage alterations. We found that dosage increase of the BOP1 gene was a frequent event, being detected in 39% of the tumors, and we show that it is associated with an increase of BOP1 mRNA. Scanning of 8q24, on which BOP1 is located, revealed that in colorectal cancers, gene dosage increase of BOP1 can be independent from that of MYC and was more frequent than that affecting MYC. Finally, transient overexpression of BOP1 in human cells increased the percentage of multipolar spindles. Together with our previous results, the present study strongly suggests that deregulation of the BOP1 pathway contributes to colorectal tumorigenesis.

PMID: 16804918 [PubMed - indexed for MEDLINE]

⊕ Publication Types, MeSH Terms, Substan

⊕ LinkOut - more resources

Related citations
Inactivation of the RRB1-Pescadillo pathway involved in ribosome biogenes [Oncogene. 2004]

Array CGH identifies distinct DNA copy number profiles of oncogenes and tum [J Mol Med. 2007]

Chromatid cohesion defects may underlie chromosome in [Proc Natl Acad Sci U S A. 2008]

Review CIN By WNT: growth pathways, mitotic control and chromosomal inst [Cell Cycle. 2006]

Review Differences between familial and sporadic forms of colorectal c [Surg Oncol. 2007]

See reviews...
See all...

Scooner @Knoesis

Welcome, Kno.e.sis   Logout

My Projects «   Dashboard   New Project *

New

Project ▼   Search: bop1 🔍   Medline / HPCO ▼   ☐ Titles Only

▲ 📁 New Project
  📁 Workbench
  ▷ 📁 Trails
  ▲ 📁 Searches
    ▤ bop1 (HPCO)
    ▤ bop1 (UMLS)

Workbench   **bop1 (HPCO)** ✕   bop1 (UMLS) ✕

Crystal structures of beta-neurexin 1 and beta-neurexin 2 ectodomains and dynamics of splice insertion sequence 4.
Presynaptic **neurexins** (NRXs) bind to postsynaptic neuroligins (NLs) to form Ca(2+)-dependent complexes that bridge neural **synapses**. beta-N present crystal structures of the **Delta isoforms** of the LNS domains from beta-NRX1 and beta-NRX2, crystallized in the presence of **Ca(2+) ions**. splice **insertion site**, with one coordinating ligand donated by a glutamic acid from an adjacent beta-NRX1 **molecule**. **NMR studies** of beta-NRX1 beta-NRX insertion sequence 4 may **function** in **roles** independent of neuroligin binding.
http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=pubmed&dopt=AbstractPlus&list_uids=18334216

Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis.
The most common form of **genomic instability** observed in **colorectal cancer is** chromosomal instability (CIN), whose molecular bases remain PES1, ORC6L, and RPL3), involved in **ribosome biogenesis**, altered chromosome segregation. To determine the contribution to colorectal tumor short fluorescent **fragments, a** sensitive method for the **detection of gene** dosage alterations. We found that dosage increase of the **BOP1 gene** BOP1 is located, revealed that in colorectal **cancers**, **gene do** more freque previous results, the present study strongly suggests that de
http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=p

Relations - BOP1 gene > related category > Gene

Relations   Expand ▼ Select ▼   Add to Workbench
▷ 📁 encode   Title
▲ 📁 related category
  ▤ Gene   ⊞ ☐ Structural organization and promoter analysis of murine heat shock tran
                ⊞ ☐ BLADE-ON-PETIOLE1 encodes a BTB/POZ domain protein required for
                ⊞ ☐ The BLADE ON PETIOLE genes act redundantly to control the growth a
                ⊞ ☐ Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigen
                ⊞ ☐ Identification by suppression subtractive hybridization of genes that are
                ⊞ ☐ Expression of the dbl proto-oncogene in Ewing's sarcomas.
                ⊞ ☐ Whole genome expression profiling reveals a significant role for immun
                ⊞ ☐ Sequence variability and candidate gene analysis in complex disease:
                ⊞ ☐ Gene expression profiling of dysplastic differentiation in cervical epithel
                ⊞ ☐ Chromosomal mapping of an alcC disruption with respect to amdA in A

Physical and functional interaction between Pes1 and Bop1 i
Molecular mechanisms of mammalian **ribosome biogenesis**   rived domin
Pes1 mutants selected by their ability to reversibly arrest the   **ells**. We sho
essential for the efficient incorporation of Pes1 into nucleolar   h Bop1 lose
ribosomal subunits.
http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=p
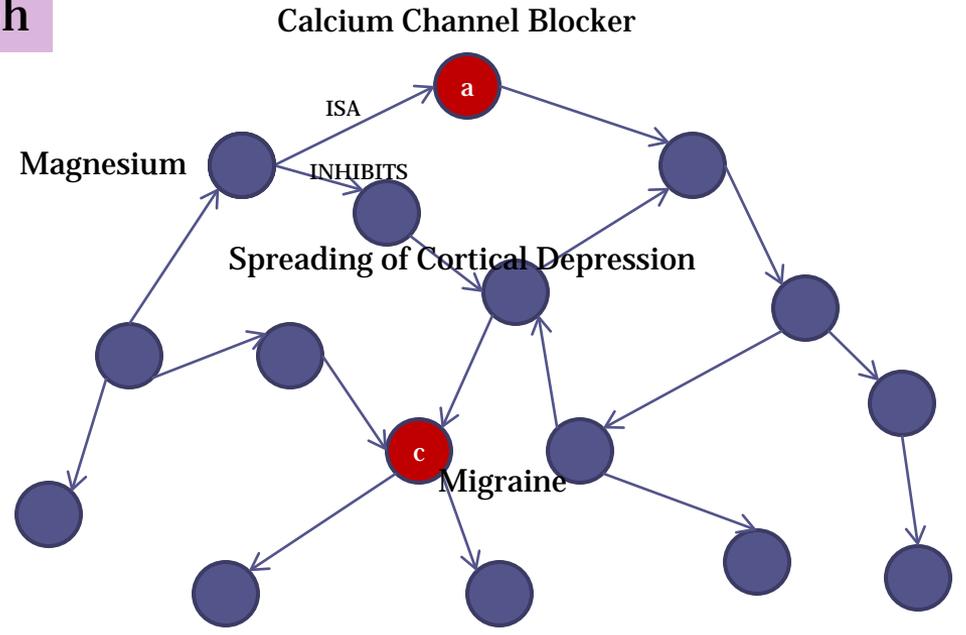
Inverse Relations   More...

Outline

- Background
- Reachability
  - Provenance Metadata
  - Knowledge Abstraction
- Approach
- Preliminary Results
- Knowledge Discovery
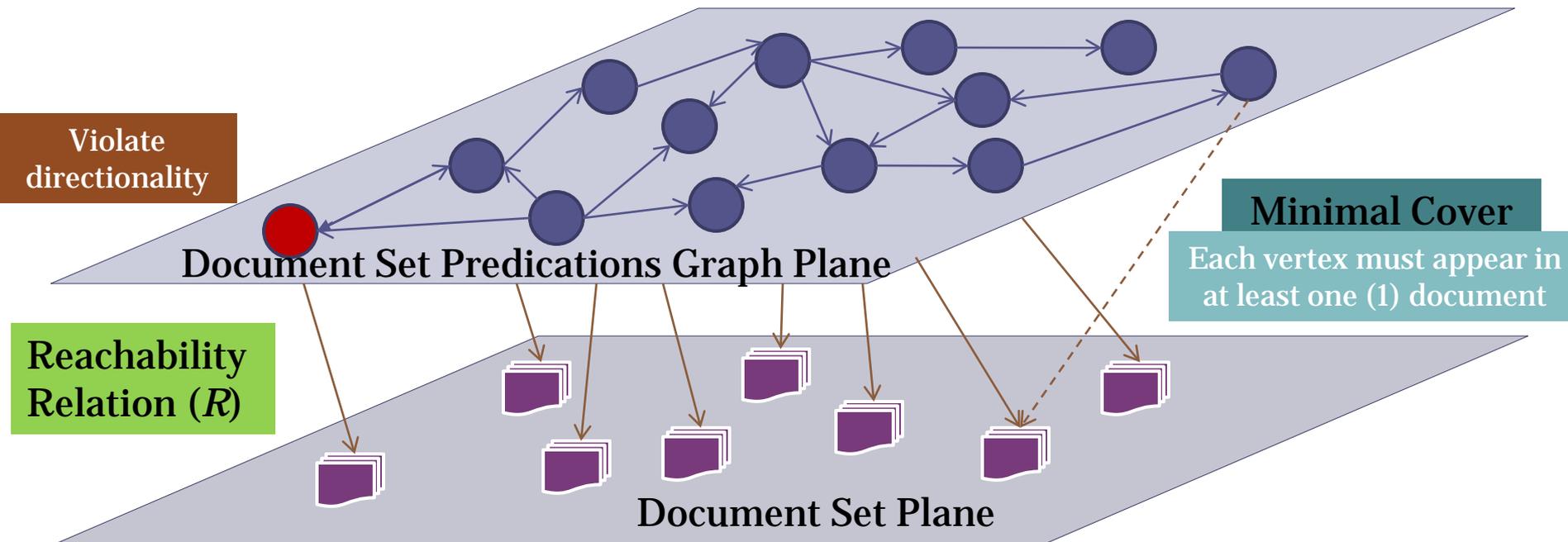- Future Work
- Conclusion

Reachability (Vertex)

"the notion of being able to get from one <u>vertex</u> in a <u>directed graph</u> to <u>some other vertex</u>"

Labeled Graph

Reachability (Documents)

"is the notion of being able to cover <u>the documents in a document set, using</u> the vertices in a <u>directed graph</u> from one <u>vertex</u> to <u>some other vertex</u>"

Violate directionality

Minimal Cover

Each vertex must appear in at least one (1) document

Document Set Predications Graph Plane

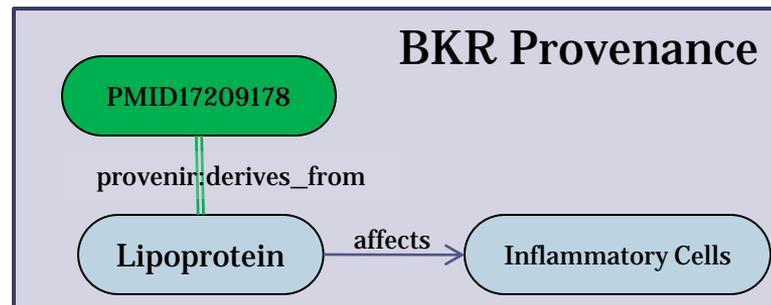Reachability Relation (R)

Document Set Plane

Outline

- Background
- Reachability
  - Provenance Metadata
  - Knowledge Abstraction
- Approach
- Preliminary Results
- Knowledge Discovery
- Future Work
- Conclusion

Provenance

- ## Meaning: origin, source of

- ## Provenance of Predications
  - ### Biomedical Knowledge Repository (BKR)
    - UMLS Metathesaurus
      - 8 million relations
    - Biomedical Literature using SemRep
      - 13.4 million predications
  - ### ProvenAnce Context Entity (PACE)[1]

BKR Provenance

PMID17209178

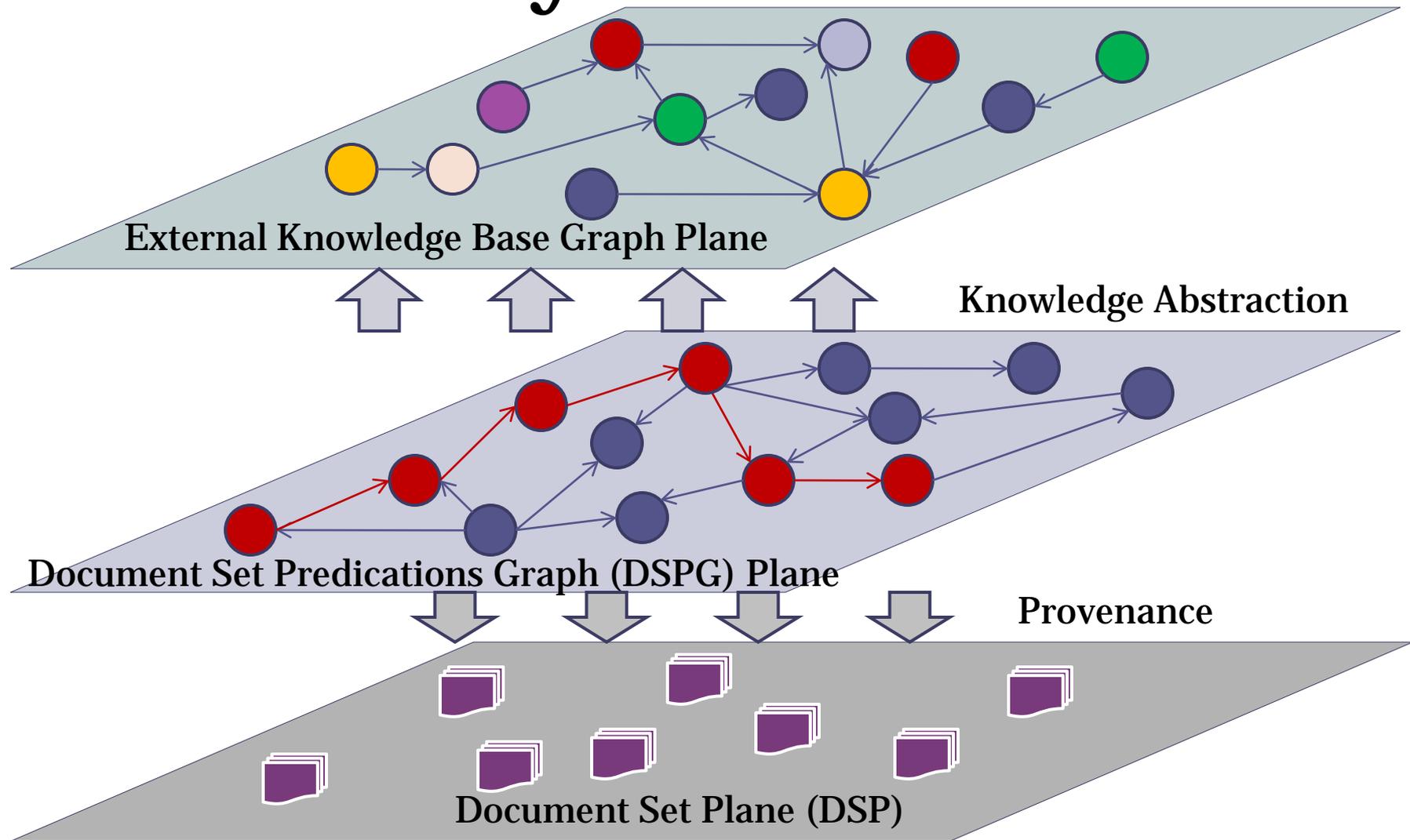provenir:derives_from

Lipoprotein →affects→ Inflammatory Cells

[1]S.S. Sahoo, O. Bodenreider, P. Hitzler, A. Sheth, K., Thirunarayan, "Provenance Context Entity (PaCE): Scalable provenance tracking for scientific RDF data.",in the 22nd International Conference on Scientific and Statistical Database Management (SSDBM) 2010

Outline

- Background
- Reachability
  - Provenance Metadata
  - Knowledge Abstraction
- Approach
- Preliminary Results
- Knowledge Discovery
- Future Work
- Conclusion

Knowledge Abstraction

- Scenarios
  - No Reachable Docs in DSP
  - No Successors in DSPG +Reachable Docs in DSP
  - No Successors in DSPG + No Reachable Docs in DSP

C0007613 – Cell physiology

coexists_with

coexists_with

C1261468 - Cell fusion

C0040682 - cell transformation

Stopping Condition (vertex)
- no successors
- no reachable docs
- no pseudo-sibling in a doc

Document Set Predications Graph (DPG) Plane

Document Set Plane (DP)

Reachability Framework

External Knowledge Base Graph Plane

Knowledge Abstraction

Document Set Predications Graph (DSPG) Plane

Provenance

Document Set Plane (DSP)

Dataset

- ## Documents (TREC 2006 Corpus)
    - 26 Questions
    - 1381 Gold Standard Documents
    - 3461 Passages

- ## Document Predications Graph
    - 2105 Vertices, 16942 Edges
    - 1141 documents
    - 240 no predications, 3 not processed

- ## External Knowledge Base Graph
    - 13 million from UMLS Metathesarus
    - 8 million from Literature using SemRep

TREC Genomics Track 2006 (Questions)

Topic ID	Question
160	What is the role of PrnP in mad cow disease?
161	What is the role of IDE in Alzheimer's disease
163	What is the role of APC (adenomatous polyposis coli) in colon cancer?
165	How do Cathepsin D (CTSD) and apolipoprotein E (ApoE) interactions contribute to Alzheimer's disease?
167	How does nucleoside diphosphate kinase (NM23) contribute to tumor progression?
168	How does BARD1 regulate BRCA1 activity?
169	How does APC (adenomatous polyposis coli) protein affect actin assembly
170	How does COP2 contribute to CFTR export from the endoplasmic reticulum?
172	How does p53 affect apoptosis?
173	How do alpha7 nicotinic receptor subunits affect ethanol metabolism?
174	How does BRCA1 ubiquitinating activity contribute to cancer?
176	How does Sec61-mediated CFTR degradation contribute to cystic fibrosis?
177	How do Bop-Pes interactions affect cell growth?
178	How do interactions between insulin-like GFs and the insulin receptor affect skin biology?
179	How do interactions between HNF4 and COUP-TF1 suppress liver function?
180	How do Ret-GDNF interactions affect liver development?
184	How do mutations in the Pes gene affect cell growth?
185	How do mutations in the hypocretin receptor 2 gene affect narcolepsy?
186	How do mutations in the Presenilin-1 gene affect Alzheimer's disease?

Graph Entry Point (GEP)

Q160: What is the role of <u>PrnP</u> in mad cow disease?

EG_5621: PRNP
C1418941: PRNP gene

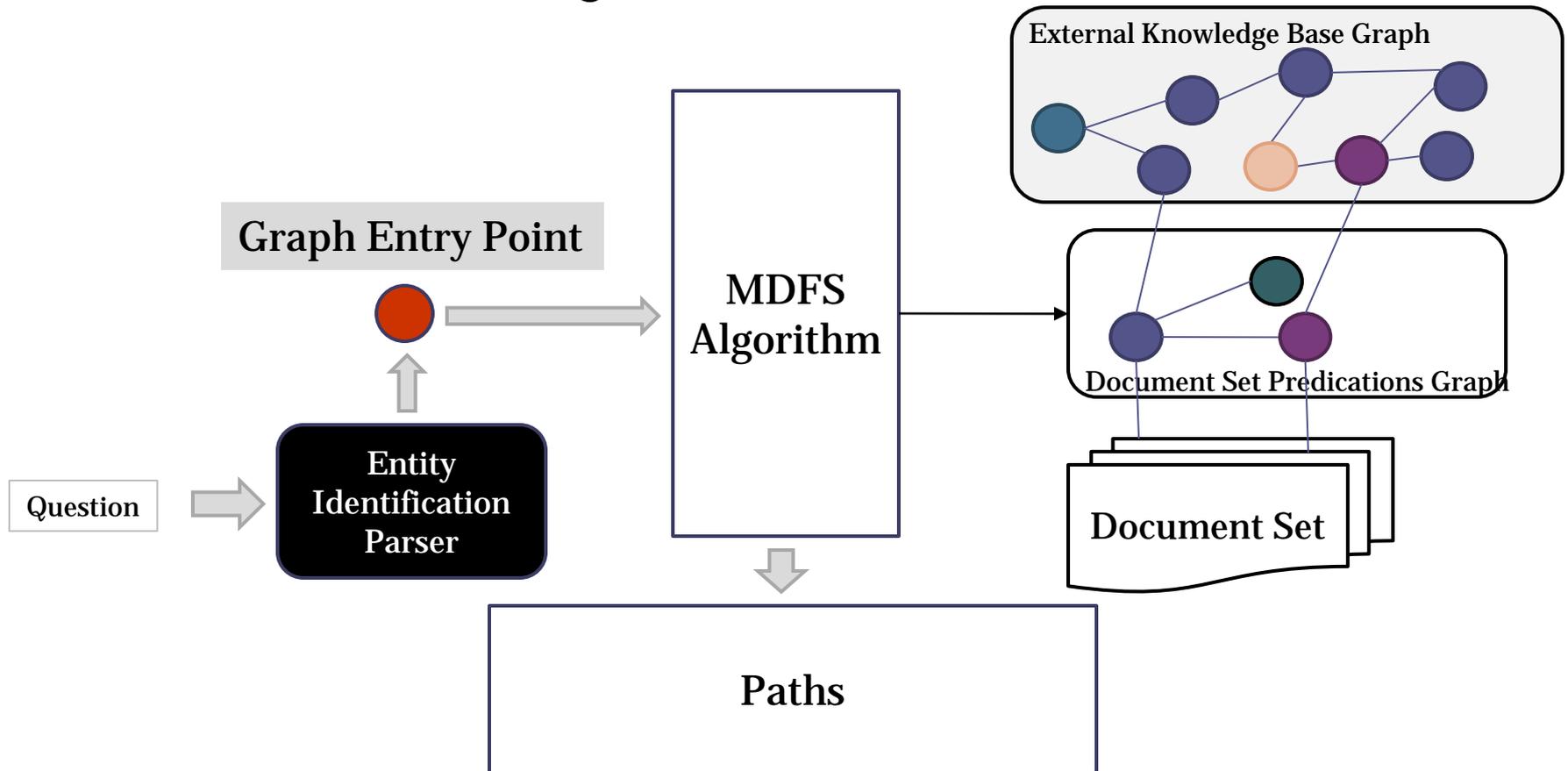Q161: What is the role of <u>IDE</u> in Alzheimer's disease

C2258633: insulin-degrading enzyme activity
C2258634: insulin-degrading neutral proteinase activity

Outline

- Background
- Reachability
    - Provenance Metadata
    - Knowledge Abstraction
- Approach
- Preliminary Results
- Knowledge Discovery
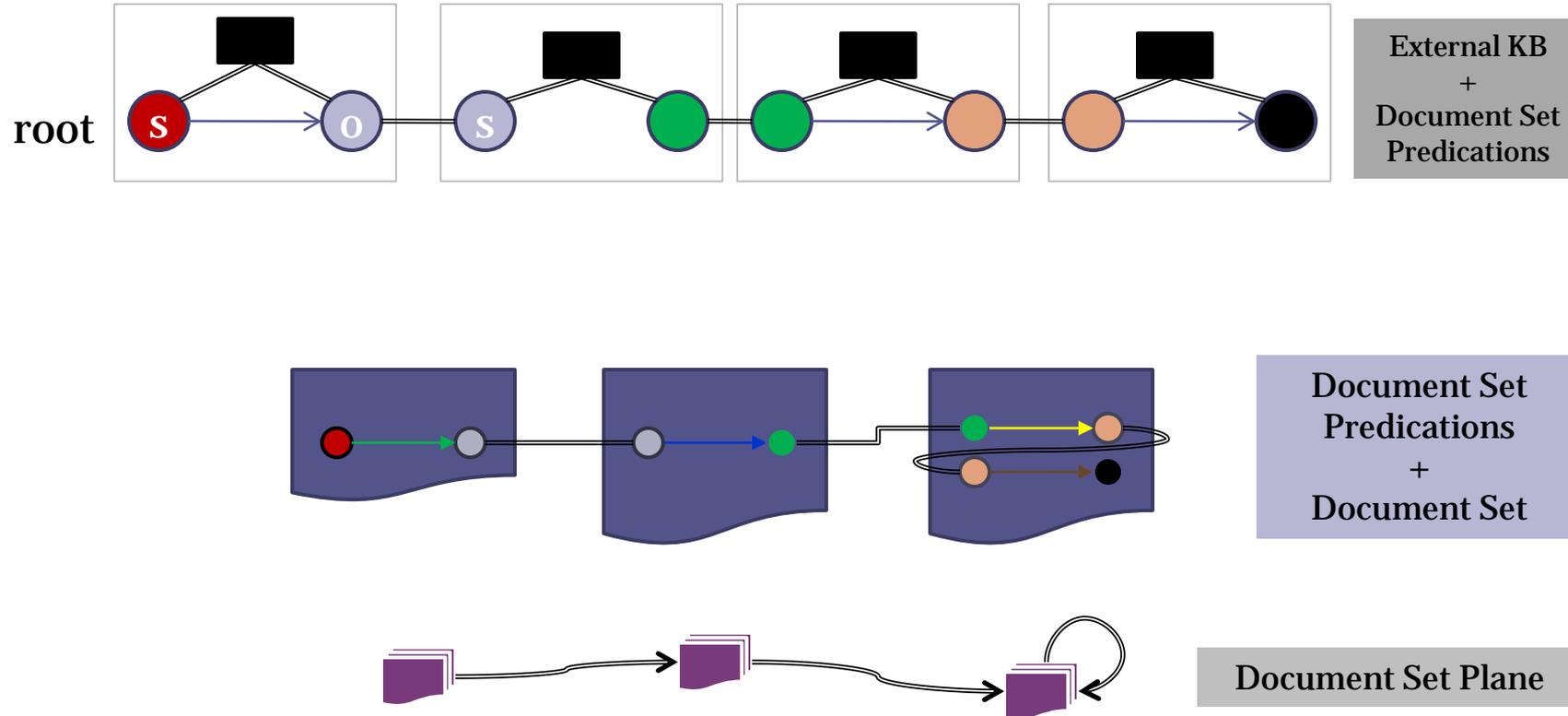- Future Work
- Conclusion

Reachability Workflow

Reachability Ratio (r_D)

$$r_D = \frac{\text{Actual Number of Documents Reached}}{\text{Total Number of Documents to be Reached}}$$

Actual: 9
Total: 10
$r_D = 9/10 = 0.9$

Reachability Path (ρ)

An ordered list of ternary relations

Path Analysis

- Inception Depth
- Minimal Maximum Reachability Path (MMRP)
- Most Relevant Subtree (MRS)
- Unconstrained Navigation

Inception Depth+MMRP

Inception Depth - Maximum number of document levels that must be explored to cover all documents

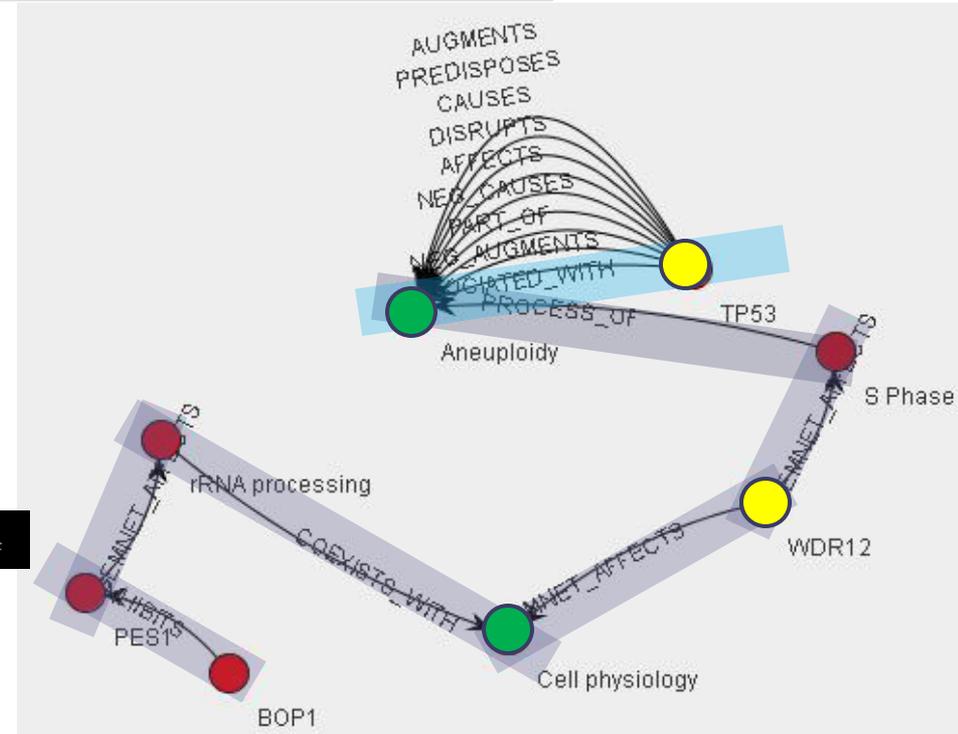**MMRP** — Paths of Minimum Length among those Paths with Maximum coverage

BOP1 → PES1 -- 16043514

PES1 → rRNA processing -- 16043514

rRNA processing → Cell physiology ← WDR12 -- 16043514

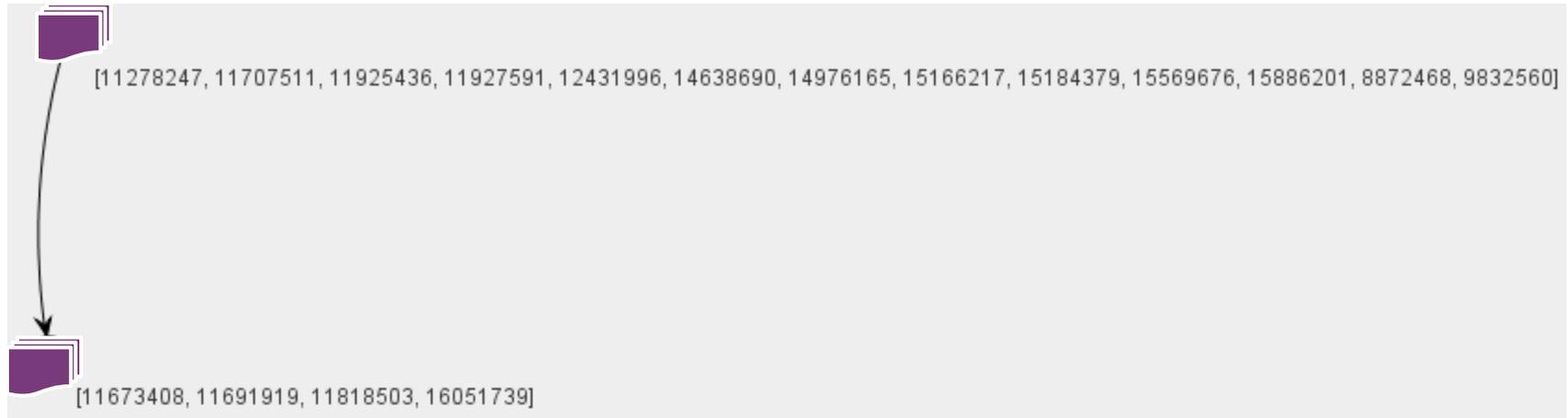WDR12 → S Phase -- 16043514

S Phase → Aneuploidy ← TP53 -- 12690111

Q177: How do Bop-Pes interactions affect cell growth?

Path Length = 5, Inception Level =2, #GS = 2, #reached = 2

Collapsed Path: BOP1 inhibits PES1 affects WDR12 affects TP53.

Inception Depth+MMRP

[11278247, 11707511, 11925436, 11927591, 12431996, 14638690, 14976165, 15166217, 15184379, 15569676, 15886201, 8872468, 9832560]

[11673408, 11691919, 11818503, 16051739]

Q174: How does BRCA1 ubiquitinating activity contribute to cancer?
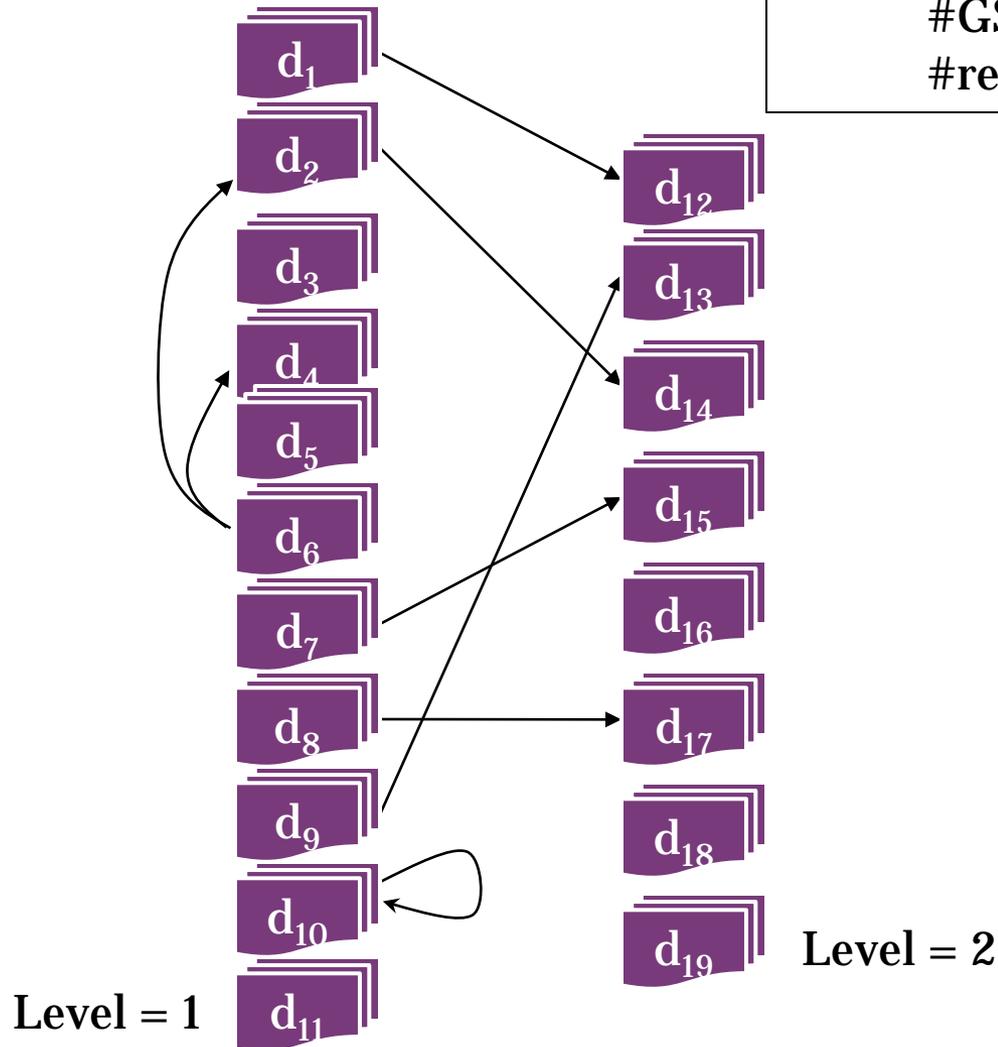
Path Length = 59
Inception Depth=2
#Gold Std Docs = 19
#reached = 17

Inception Depth

Q174: Path Length = 59
        Inception Depth = 2
        #GS = 19
        #reached = 17

d_1

d_2

d_3

d_4

d_5

d_6

d_7

d_8

d_9

d_{10}

d_{11}

d_{12}

d_{13}

d_{14}

d_{15}

d_{16}

d_{17}

d_{18}

d_{19}

Level = 1

Level = 2

Outline

- Background
- Reachability
  - Provenance Metadata
  - Knowledge Abstraction
- Approach
- Preliminary Results
- Knowledge Discovery
- Future Work
- Conclusion

Path Analysis

Topic ID	Graph Entry Point (GEP)	#reached/#GS docs	Reachability Ratio (r_D)	Inception Depth (δ)	MMRP (ρ_{mm})
165	C0007427/CATHEPSIN D	9/10	0.90	2	6
168	C1332381/BARD1 gene	30/43	0.69	9	65
170	EG_1080/CFTR (try C0538130/COP23)	3/4	0.75	2	21
174	EG_672/BRCA1	17/19	0.89	2	59
176	EG_1080/CFTR	9/11	0.82	3	12
177	EG_23246/BOP1	2/2	1.0	2	5
178	C0021665/Insulin-Like Growth Factor I	3/3	1.0	3	7
179	EG_3172/HNF4A	12/12	1.0	4	32
184	EG_23481/PES1	2/2	1.0	2	4
185	C1113688/orexins	12/13	0.92	3	13

Outline

- Background
- Reachability
  - Provenance Metadata
  - Knowledge Abstraction
- Approach
- Preliminary Results
- Knowledge Discovery
- Future Work
- Conclusion

Most Relevant Subtree

Reachability Relation
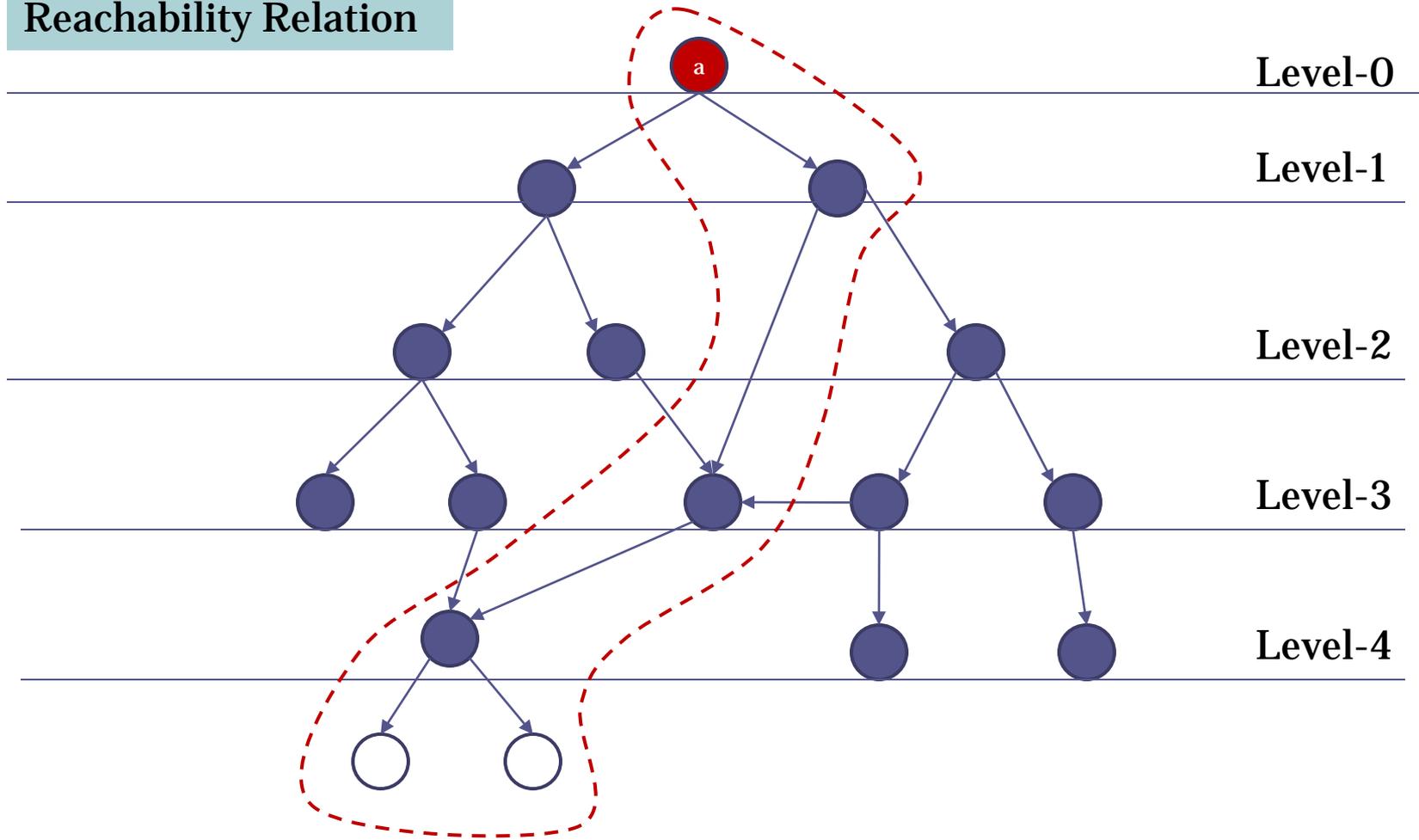
Unconstrained Traversal

Q177: How do Bop-Pes interactions affect cell growth?

MRS: [BOP1, PES1, rRNA processing, _Cell physiology_, WDR12, S Phase]

Level-0: 1 entity

BOP1/GENE

Level-1: 3 entities

PES1/GENE

Level-2: 137 entities

rRNA processing/Genetic Function

Level-3: 5956 entities

Cell physiology/Cell Function

Level-4: 95080 entities

WDR12/GENE

Future Work

- • **Complete Reverse Engineering**

- • **Pattern Learning**
  - ▪ Statistical Analysis
  - ▪ Machine Learning

- • **Extensive Knowledge Abstraction**
  - ▪ Associative
  - ▪ Hierarchical
  - ▪ Hybrid

- • **Target Venue**
  - ▪ WWW 2011, Hyderabad, India.
  - ▪ Deadline October 29, 2010

Conclusion

- Reachability
  - Provenance
  - Knowledge Abstraction
  - Path Analysis

- Knowledge Discovery
  - Reverse Engineering
  - Pattern Filtering

Acknowledgements