



CCI, UNC-Charlotte



CS Dept., PJAiT, Warsaw, Poland

Knowledge-Graph Construction

Zbigniew W. Ras

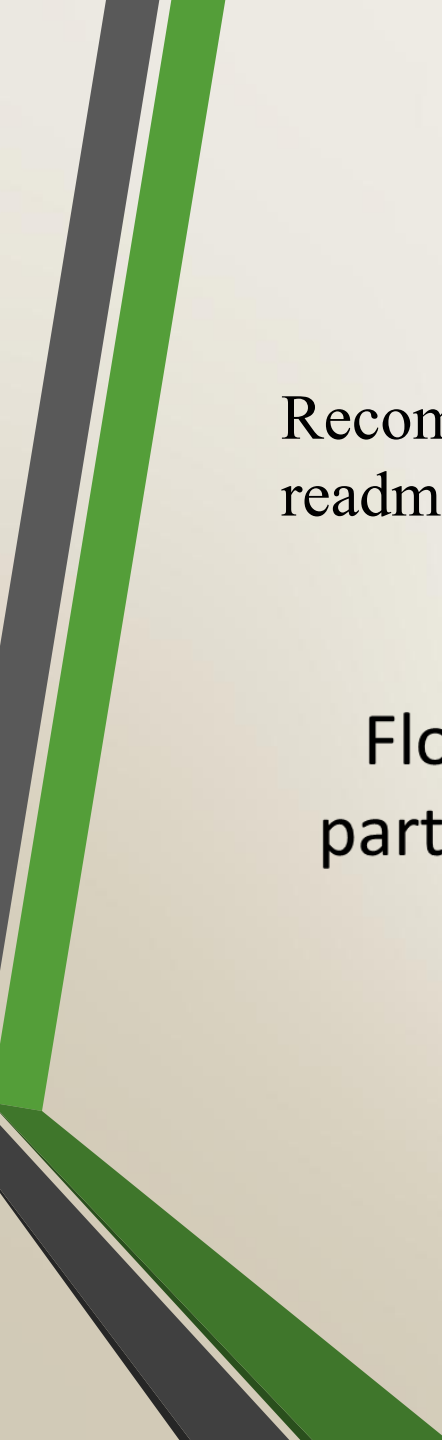
University of North Carolina at Charlotte

&

Polish-Japanese Academy of Information Technology

Sponsored by 





Recommendations/interventions leading to
readmission reduction

Florida State Inpatient Databases (SID),
part of the Healthcare Cost and Utilization
Project (HCUP)

HCUP Dataset

- Part of the Healthcare Cost and Utilization Project (HCUP).
- A total of over 7.8 million visit discharges for over 3.6 million patients in Florida..
- Each patient can exhibit up to 31 diagnoses and up to 31 procedures. (The first procedure is considered the main procedure).
- Diagnoses/Procedures codes:
 - **International Classification of Diseases**, Ninth Revision, Clinical Modification (ICD-9-CM)
 - **Clinical Classifications Software** (CCS)

Diagnoses: ~15k (ICD-9-CM), ~300 (CCS)

Procedure codes: ~4k (ICD-9-CM), 231 (CCS)

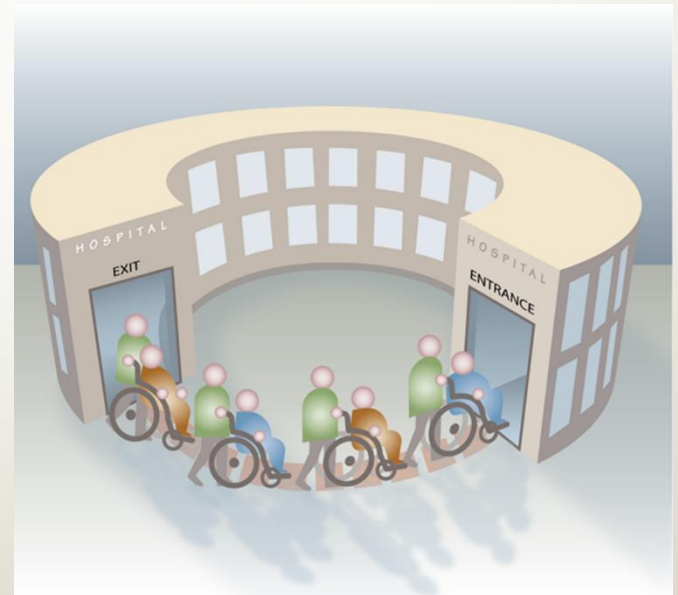
CCS CODING VS ICD-9-CM CODING

- CCS coding is a summarized version of ICD-9-CM coding; it means CCS groups a set of diagnoses into a more general category.
- Using ICD-9-CM code, we have more descriptive and specific diagnoses and procedure codes

	NUMBER OF CCS CODES	NUMBER OF ICD-9-CM CODES
PROCEDURE CODES	231	3,948
DIAGNOSES CODES	300	15,072

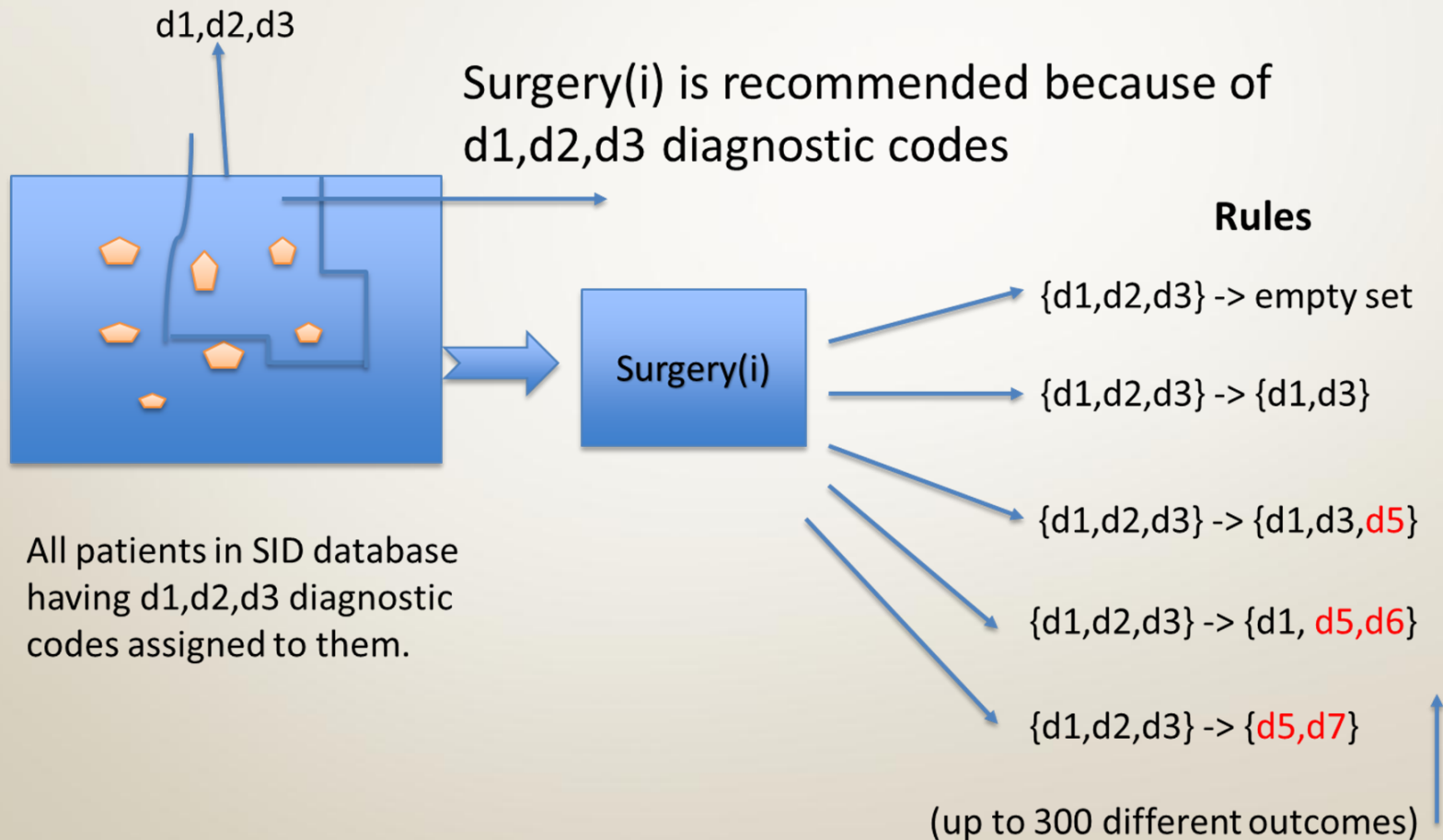
Hospital Readmissions

Hospital Readmission is defined as a re-hospitalization of a patient after being discharged from a hospital within a short period of time. The period in average is 30 days

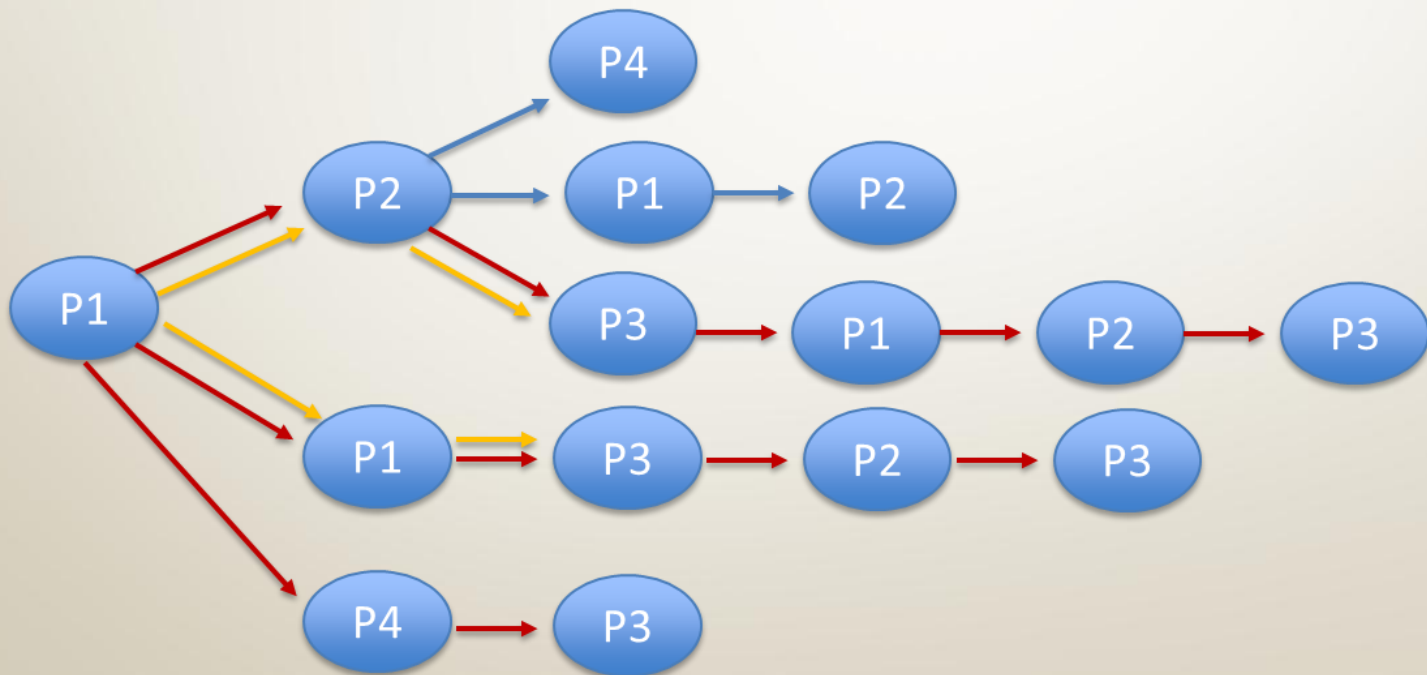


Hospital Readmissions

Where is the problem?



- Procedure paths: sequence of procedures that a given patient undertakes to reach a desired treatment.
- We study the problem of predicting procedure paths and providing a foundation for transitioning from one procedure path to another



Example: Procedure Code 78 (Colorectal Resection)

41,753 patients started with this procedure.

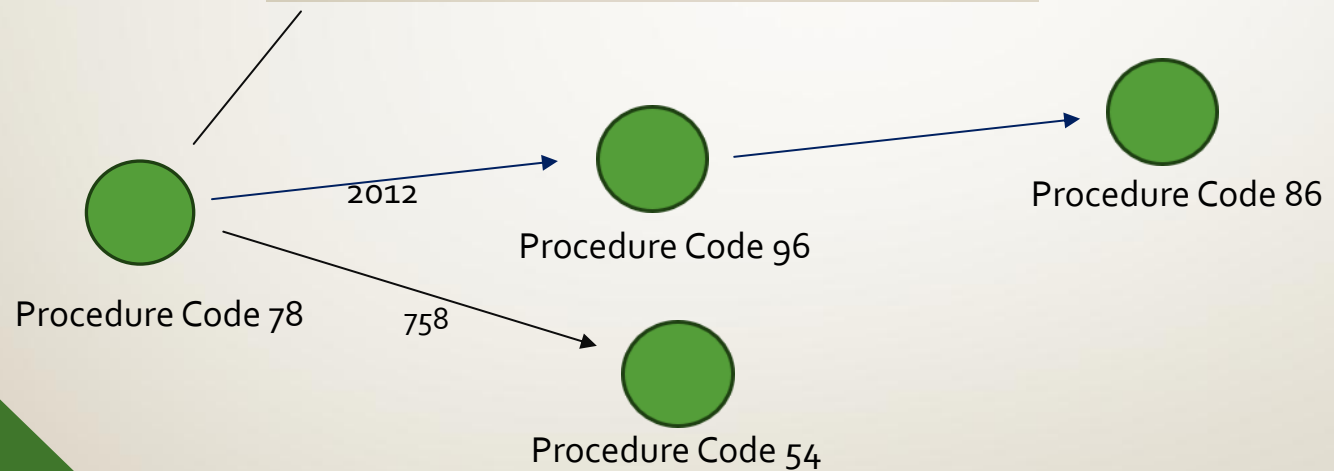
6,774 unique procedure paths.

Most probable 2-element paths

Procedure code	Explanation	Probability	Support
78, 96	96:Other OR lower GI therapeutic procedures	4.8 %	2012
78, 78	78	2.5 %	1074
78, 54	54:Other vascular catheterization; not heart	1.8 %	758
78, 222	222: Blood transfusion	1.6 %	698
78, 90	90: Excision; lysis peritoneal adhesions	1.6 %	656

Knowledge Graph Construction

Procedure code	Explanation	Probability	Support
78, 96	96:Other OR lower GI therapeutic procedures	4.8 %	2012
78, 78	78	2.5 %	1074
78, 54	54:Other vascular catheterization; not heart	1.8 %	758
78, 222	222: Blood transfusion	1.6 %	698
78, 90	90: Excision; lysis peritoneal adhesions	1.6 %	656



Example: Procedure Code 78 (Colorectal Resection)

41,753 patients started with this procedure.

6,774 unique procedure paths.

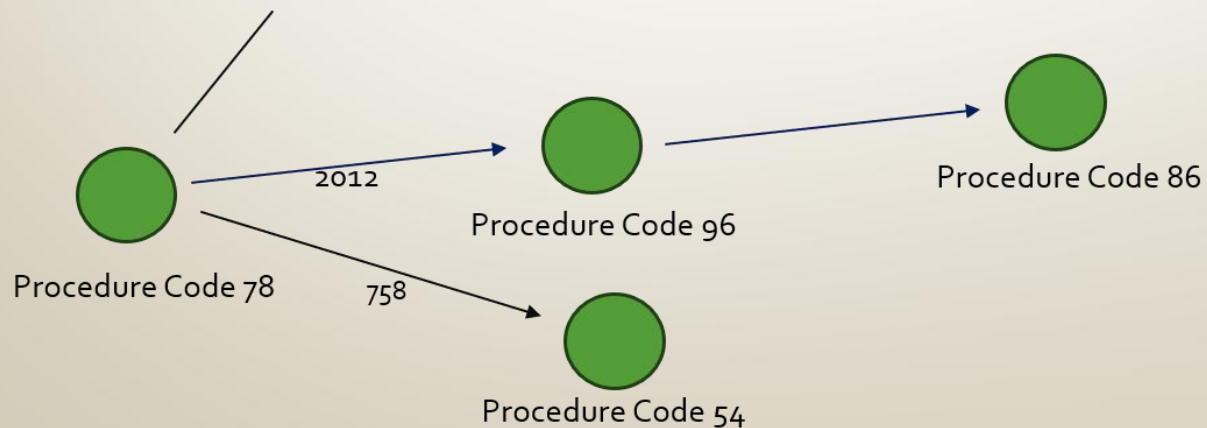
Most probable 3-element paths

Procedure code	Explanation	Probability	Support
78, 224, 224	224: Cancer chemotherapy	0.23 %	94
78, 96, 96	96:Other OR lower GI therapeutic procedures	0.20 %	82
78, 78, 96	78, 96	0.18 %	76
78, 222, 222	222: Blood transfusion	0.15 %	63
78, 96, 86	90: Other hernia repair	0.14 %	59
...			
78, 78, 78	78	0.045 %	19

Procedure code	Explanation	Probability	Support
78, 224, 224	224: Cancer chemotherapy	0.23 %	94
78, 96, 96	96:Other OR lower GI therapeutic procedures	0.20 %	82
78, 78, 96	78, 96	0.18 %	76
78, 222, 222	222: Blood transfusion	0.15 %	63
78, 96, 86	90: Other hernia repair	0.14 %	59
...			
78, 78, 78	78	0.045 %	19

Knowledge Graph Construction

Procedure code	Explanation	Probability	Support
78, 96	96:Other OR lower GI therapeutic procedures	4.8 %	2012
78, 78	78	2.5 %	1074
78, 54	54:Other vascular catheterization; not heart	1.8 %	758
78, 222	222: Blood transfusion	1.6 %	698
78, 90	90: Excision; lysis peritoneal adhesions	1.6 %	656



Procedure 78: Longest Path

Length 22 (1 path):

78(Colorectal resection), 223(Enteral and parenteral nutrition), 223(Enteral and parenteral nutrition), 70(Upper gastrointestinal endoscopy; biopsy), 75(Small bowel resection), 223(Enteral and parenteral nutrition), 172(Skin graft), 223(Enteral and parenteral nutrition), 223(Enteral and parenteral nutrition), 223(Enteral and parenteral nutrition), 223(Enteral and parenteral nutrition), 54(Other vascular catheterization; not heart), 223(Enteral and parenteral nutrition), 54(Other vascular catheterization; not heart), 70(Upper gastrointestinal endoscopy; biopsy), 54(Other vascular catheterization; not heart), 223(Enteral and parenteral nutrition), 54(Other vascular catheterization; not heart), 54(Other vascular catheterization; not heart), 34(Tracheostomy; temporary and permanent), 54(Other vascular catheterization; not heart),

Probability is 2.396070198582124E37, Support: 1

Procedure 78: Length 21 (2 paths):

78(Colorectal resection), 221(Nasogastric tube), 96(Other OR lower GI therapeutic procedures), 231(Other therapeutic procedures), 231(Other therapeutic procedures), 221(Nasogastric tube), 221(Nasogastric tube), 78(Colorectal resection), 78(Colorectal resection), 223(Enteral and parenteral nutrition), 223(Enteral and parenteral nutrition), 221(Nasogastric tube), 222(Blood transfusion), 54(Other vascular catheterization; not heart), 223(Enteral and parenteral nutrition), 223(Enteral and parenteral nutrition), 223(Enteral and parenteral nutrition), 174(Other non-OR therapeutic procedures on skin and breast), 174(Other non-OR therapeutic procedures on skin and breast), 223(Enteral and parenteral nutrition), 223(Enteral and parenteral nutrition),

Probability is 2.3960703494154103E35, Support: 1

78(Colorectal resection), 222(Blood transfusion), 58(Hemodialysis), 88(Abdominal paracentesis), 58(Hemodialysis), 58(Hemodialysis), 88(Abdominal paracentesis), 58(Hemodialysis), 58(Hemodialysis), 58(Hemodialysis), 58(Hemodialysis), 58(Hemodialysis), 88(Abdominal paracentesis), 58(Hemodialysis), 47(Diagnostic cardiac catheterization; coronary arteriography), 34(Tracheostomy; temporary and permanent), 216(Respiratory intubation and mechanical ventilation), 58(Hemodialysis), 40(Other diagnostic procedures of respiratory tract and mediastinum), 58(Hemodialysis), 222(Blood transfusion),

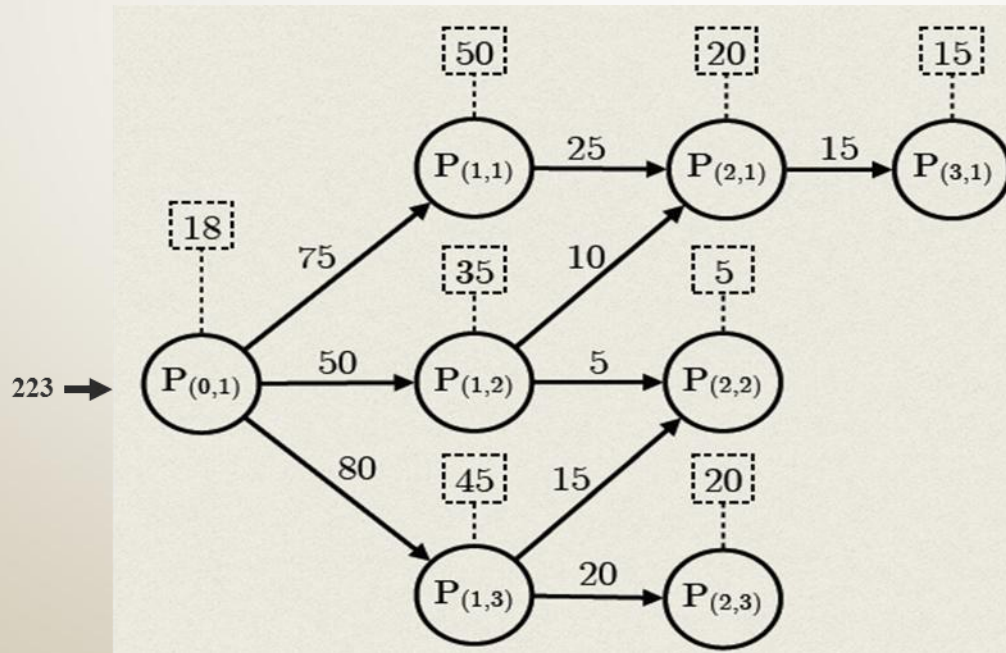
Probability is 2.3960705057112934E35, Support: 1



Knowledge Graph

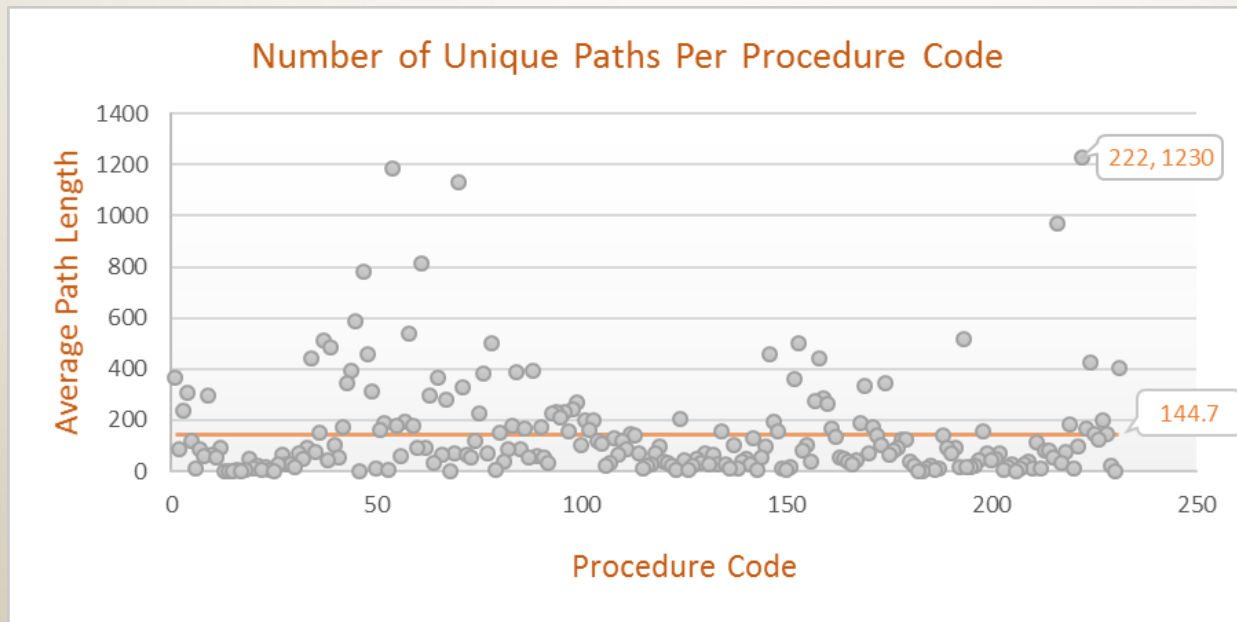
Knowledge Graph

Knowledge graph for some procedure p is defined as a graph representing all possible procedure paths extracted from our dataset for patients who underwent procedure p as their first procedure



Information Extracted from Knowledge Graph

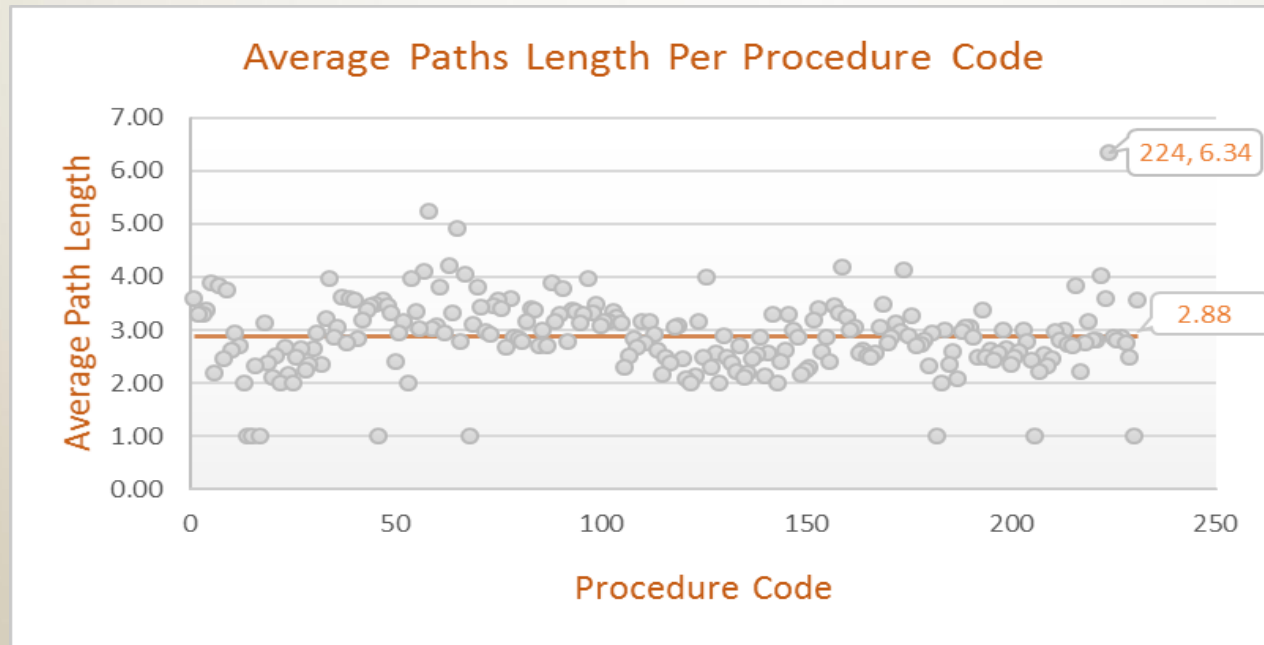
- The number of all procedure paths is extremely high.
- It does not exist a single universal course of treatment that patients typically follow to reach the desired state.



222 : Blood Transfusion

Information Extracted from Knowledge Graph

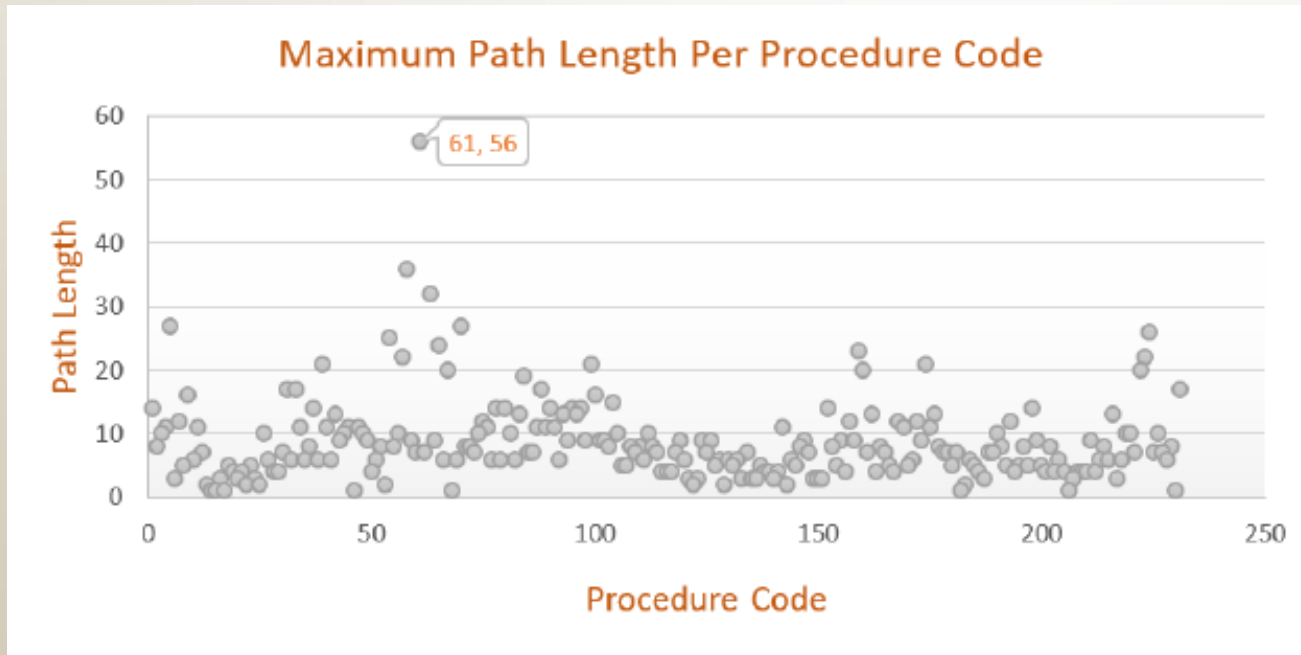
- The length of the procedure path is an indicator of the number of readmissions during the course of treatment.
- Path length (Number of readmission) = Number of nodes (procedures)-1



224 : Cancer chemotherapy

Information Extracted from Knowledge Graph

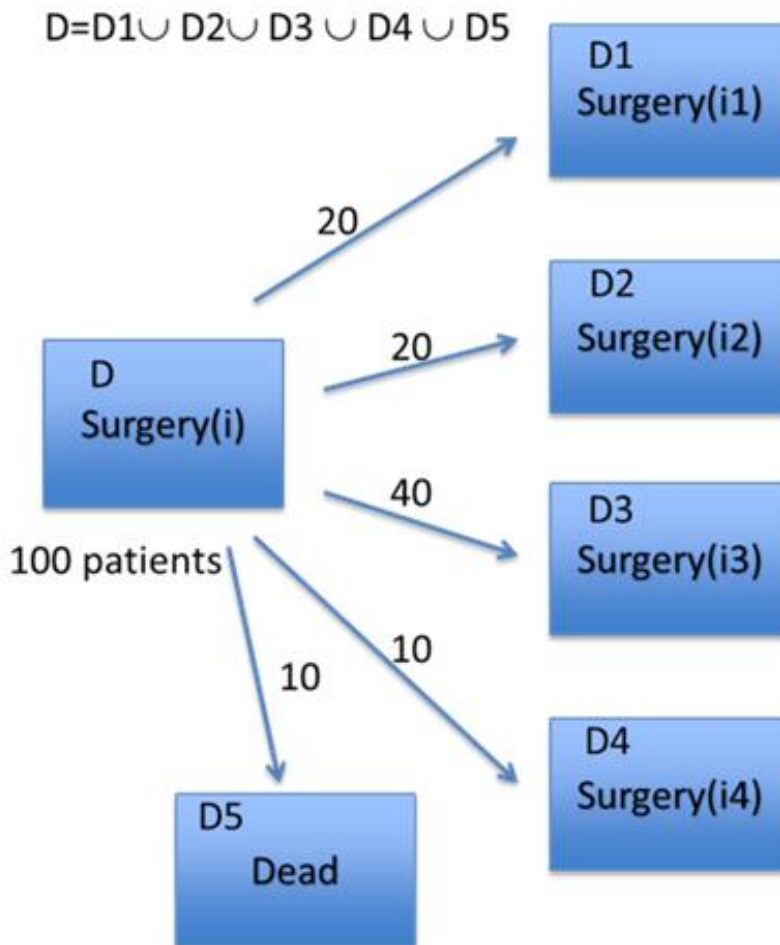
- Knowing that patient will undergo procedure 58 (Hemodialysis) as the initial procedure allows us to anticipate what could be the maximum and the average number of possible readmissions in general, which are 36 and 4.2 respectively.



61 : procedures on vessels other than head and neck



Knowledge Graph Personalization



System S	Side effects	Next Procedure
# Patients	A	D
10	A1	D2
10	A1	D1
10	A1	D1
5	A2	D2
10	A2	D4
10	A3	D5
5	A4	D2
10	A4	D3
30	A5	D3

$$\text{Entropy}(S,A) = 3/10[2/3[\log(2/3)] + 1/3[\log(1/3)]] + 15/100[1/3\log(1/3) + 2/3\log(2/3) + \dots] +$$

Patients' Clustering

Clustering criteria:

Race

Gender

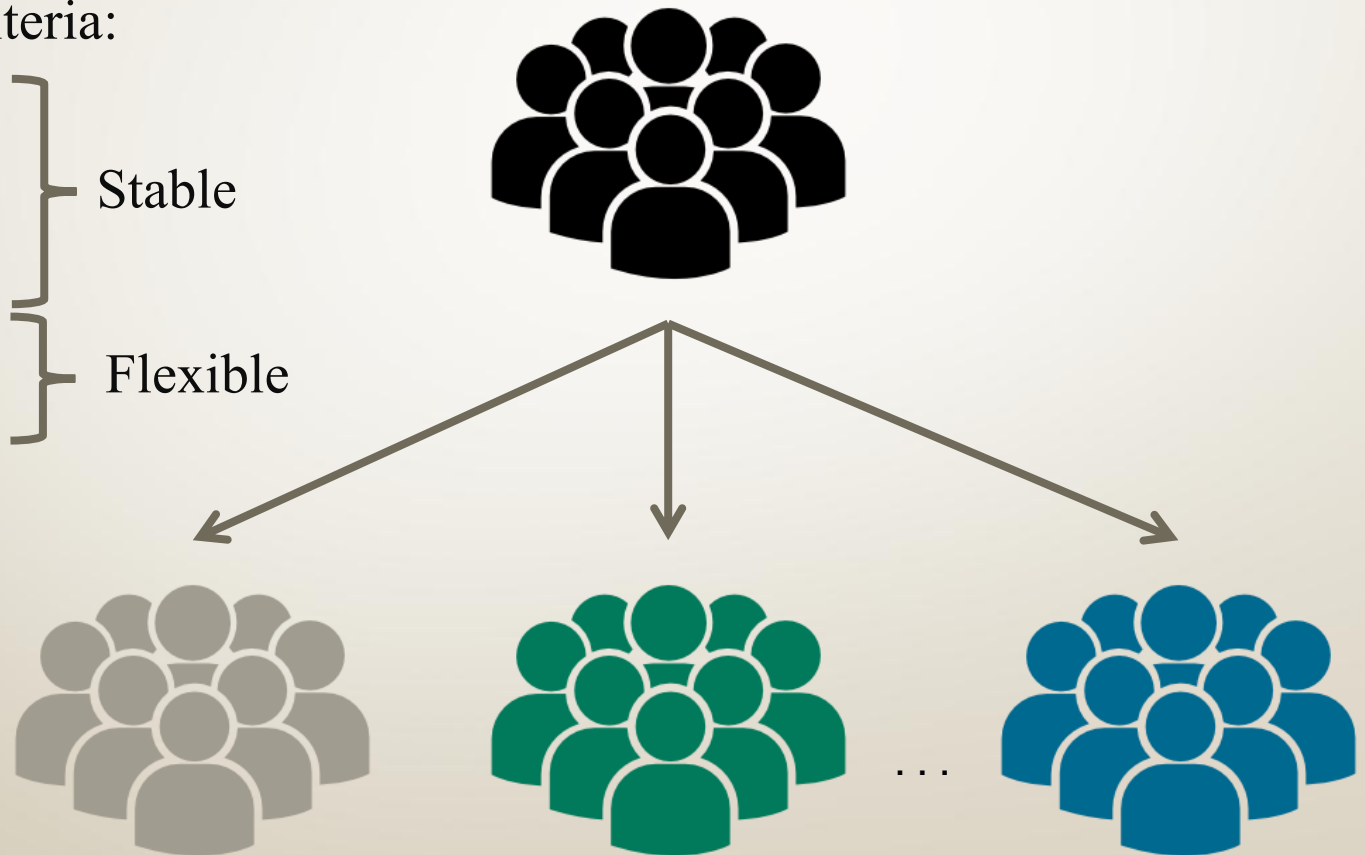
Age

Diagnoses

Procedures

Stable

Flexible



Stable Features Clusters

Race	Code
White	1
Black	2
Hispanic	3
Asian or Pacific Islander	4
Native American	5
Other	6

Age	Range
Child	[0 – 17]
Young	[18 – 44]
Mature	[45 – 65]
Old	[65 –]

•
•
•

White, Child	Black, Child	Hispanic, Child	Asian, Child	Native, Child	Other, Child
White, Young	Black, Young	Hispanic, Young	Asian, Young	Native, Young	Other, Young
White, Mature	Black, Mature	Hispanic, Mature	Asian, Mature	Native, Mature	Other, Mature
White, Old	Black, Old	Hispanic, Old	Asian, Old	Native, Old	Other, Old

Personalization Using Stable Features

Procedure Code	Number of Patients	Entropy			
		No Clustering	Clustering based on		
			Race only	Age only	Race and Age
48	15139	5.657	5.608	5.595	5.518
44	7705	5.684	5.604	5.615	5.495
58	9187	3.914	3.855	3.853	3.748
158	17129	5.29	5.249	5.181	5.103
39	8717	4.993	4.927	4.866	4.753

Slight decrease in entropy

48: Insertion; revision; replacement; removal of cardiac pacemaker or cardioverter/defibrillator,
44: Coronary artery bypass graft (CABG)
58: Hemodialysis, 158: Spinal fusion, 39: Incision of pleura; thoracentesis; chest drainage

Exact Matching Clustering-Example

- Procedure: 158 (Spinal fusion)

Set of Diagnoses	Number of Patients	Entropy
{205}	502	3.015
{205, 98}	128	2.752
{205, 663}	86	2.543
{205, 209}	67	2.798
{205, 211}	51	2.510

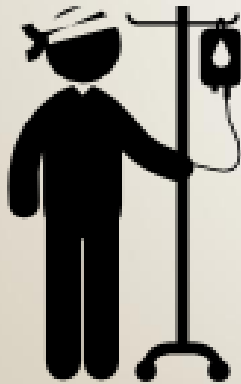
- 205: Spondylosis; intervertebral disc disorders; other back problems
 - 98: Essential hypertension
 - 663: Screening and history of mental health and substance abuse codes
 - 209: Other acquired deformities
 - 211: Other connective tissue disease
- Entropy of the whole system before personalization is **3.667**
 - Weighted entropy for all exact matching clusters that have size above threshold 50 is **2.892**

Rough Clustering of Patients



- Patients are clustered based on diagnoses.
- We define two sets of diagnoses that will determine whether a patient belongs to a particular cluster or not:
 - *Included set*: describes the set of diagnoses that a patient has to exhibit to belong to a given cluster.
 - *Excluded set*: describes the set of diagnoses that a patient cannot exhibit to belong to a given cluster.

Rough Clustering - Example



$\{2391, 2749\}$
 $\{-2859, -3723\}$



$\{2391, 2749\}$
 $\{-2859\}$

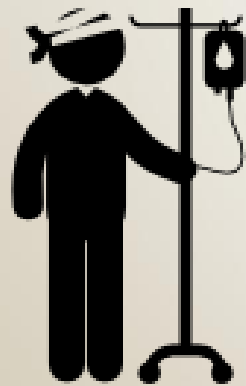


$\{2749\}$
 $\{-2859, -3723\}$



$\{\}$
 $\{-2859, -3723, -2749\}$

Rough Clustering - Example



Level 3



$\{2391, 2749\}$
 $\{-2859, -3723\}$



$\{2391, 2749\}$
 $\{-2859\}$

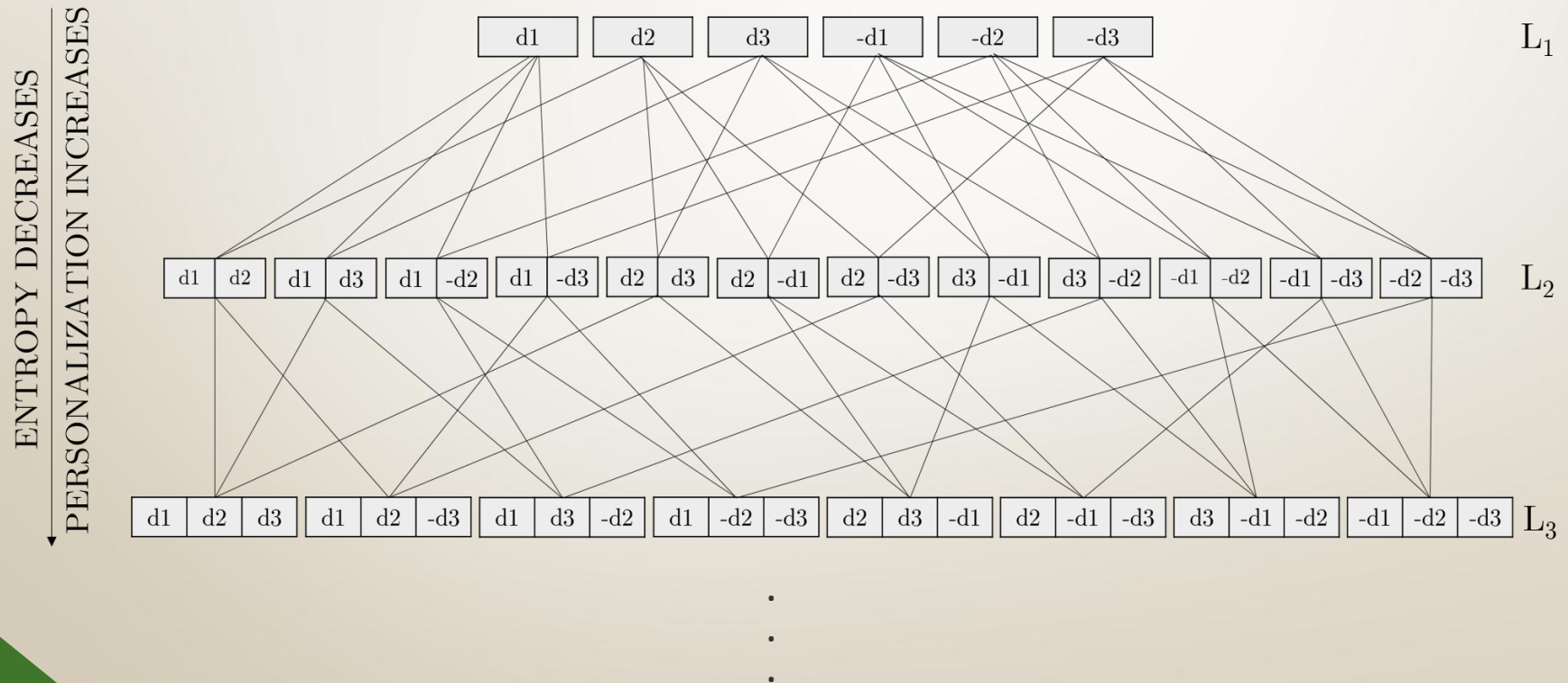


$\{2749\}$
 $\{-2859, -3723\}$



$\{\}$
 $\{-2859, -3723, -2749\}$

Rough Clustering - Implementation



Rough Clustering - Filtering

- If all patients have a common diagnosis, then this diagnosis will not play any role in determining the state for which patients end up in.
- We only considered diagnostic codes that lie within a specific range. *For example*, when we choose the allowed range to be between 20% and 80%, this means that we only consider diagnoses for which the number of patients that exhibit that diagnosis is between 20% and 80%. **It means only these diagnoses are considered in clusters construction.**
- This procedure is also applied to the excluded set, meaning that we only consider diagnoses that were missing from 20% to 80% of the total number of patients.

Procedure 158
Spinal Fusion

Number of clusters increases

ENTROPY INCREASES
↓
NUMBER OF CLUSTERS
INCREASES

	Range					
	20% to 80%		10% to 90%		5% to 95%	
	# of clusters	Entropy	# of clusters	Entropy	# of clusters	Entropy
1-element clusters	14	5.105	36	5.109	66	5.117
2-element clusters	37	5.051	379	5.028	1532	5.041
3-element clusters	50	4.988	2097	4.958	19167	4.971
4-element clusters	44	4.916	6969	4.905	155028	4.91

Rough Clustering - Filtering

- Rough clustering generates a large number of clusters
 - Number of diagnoses = 283
 - Number of 1-element sets = 566
 - Number of 2-element sets = 159,612
 - Number of 3-element sets = 29,980,454

Table 1. Number of clusters and entropy for different element sets and different ranges for **procedure 158** (spinal fusion)

Number of clusters increases

ENTROPY INCREASES

NUMBER OF CLUSTERS
INCREASES

	Range					
	20% to 80%		10% to 90%		5% to 95%	
	# of clusters	Entropy	# of clusters	Entropy	# of clusters	Entropy
1-element clusters	14	5.105	36	5.109	66	5.117
2-element clusters	37	5.051	379	5.028	1532	5.041
3-element clusters	50	4.988	2097	4.958	19167	4.971
4-element clusters	44	4.916	6969	4.905	155028	4.91

Rough Clustering – Knowledge Graph Personalization

