COMMUNITY DETECTION AND COURSE RECOMMENDATION IN ACADEMIC NETWORKS

Jobin Wilson, Anupama Ray, Shraddha Chaudhary

Department of Electrical Engineering

Indian Institute of Technology, Delhi

PROBLEM STATEMENT

- Academic community detection
  - Project/study groups

- Course recommendation
  - Similar Courses (content based)
  - Course history based Recommendations

Raw Data :

- Course List
- Prospectus

COURSES OF STUDY 2011-2012

EEL708 Information Retrieval
3 credits (3-0-0)
Pre-requisites: CSL201
Introduction – What is IR; Applications and significance; retrieval evaluation; Query Modeling and Query Languages; Indexing and Searching Text; Multimedia IR: Models, indexing , searching; User Interfaces and Visualisation; Distributed IR; Web Search Engines; Digital Libraries.

EEL709 Pattern Recognition
3 credits (3-0-0)
Pre-requisites: MAL250
Introduction – What is Pattern Recognition? Applications and Relation with other fields like Data Mining, Information Retrieval, etc.; Linear Discriminant Functions and its Applications; Bayesian Decision Theory; Maximum-Likelihood and Bayesian Parameter Estimation; Component Analysis, Expectation Maximisation, Hidden Markov Model; Non-parametric Techniques; Nearest Neighbour, K-NN; Non-metric Methods; Decision Trees, ID3, Grammar based Methods; Neural Network Based Approaches; Introduction to Fuzzy Logic Based Techniques; Support Vector Machine; Applications.

User groups and mailing lists of courses

AMD310	AMD812	AMD813	AMD814	AMD897
AML110A	AML110B	AML120	AML130	AML150
AML300	AML310	AML360	AML410	AML700

User Group: EEL709
email: EEL709@courses.iitd.ac.in

EEL709	
crf113637	Baiju M Nair
cs1090175	Abhishek Kumar
cs1090181	Ankita
cs1090208	Pritesh Kumar

PRE PROCESSING AND CONSOLIDATION OF DATA

course_id	title	credits	contents	name	name	entry_id
CSL864	Special Topics in Artificial Intelligence	3	Pre-requisites: CSL433 Under this topic one of the following areas...	K K Biswas	Jobin Wilson	eey127507
EEL709	Pattern Recognition	3	Pre-requisites: MAL250 Introduction - What is Pattern Recognition...	Sumeet Agarwal	Jobin Wilson	eey127507
EEL781	Neural Networks	3	Pre-requisites: EEL205 Overlaps with: MAL720 Introduction to Biol...	Jayadeva	Jobin Wilson	eey127507
EEL851	Special Topics in Computers - I	3	Pre-requisites: CSL201 and EC 120 Topics of current interest relat...	Santanu Chaudhury	Jobin Wilson	eey127507
MAL704	Numerical Optimization	3	Algorithms and complexity, the classes P and NP. NP complete pr...	Aparna Mehra	Jobin Wilson	eey127507

course_id	credits	lpt	title	contents
CEL743	3	(3-0-0)	Economic Aspects of Water Resources Devlopment	Data requirements and survey: topographical, geological, hyd...
CSL333	4	(3-0-2)	Artificial Intelligence	Pre-requisites: CSL201 Overlaps with: EEL758 Problem solvi...
CSL671	4	(3-0-2)	Artificial Intelligence	Problem solving, search techniques, control strategies, gam...
CSL864	3	(3-0-0)	Special Topics in Artificial Intelligence	Pre-requisites: CSL433 Under this topic one of the following ...
EEL365	3	(3-0-0)	Intelligent Control	Pre-requisites: EEL301 Conventional control review, Feedbac...
EEL462	3	(3-0-0)	Identification and Adaptive Control	Pre-requisites: EEL301 Review of stochastic process. Model...
EEL706	3	(3-0-0)	Soft Computing	Pre-requisites: MAL220 / MAL250 Overlaps with: MAL717, MA...
EEL781	3	(3-0-0)	Neural Networks	Pre-requisites: EEL205 Overlaps with: MAL720 Introduction to...
EEV835	1	(1-0-0)	Special Module in Machine Learning	Special module that focuses on special topics, Development ...
HUL261	4	(3-1-0)	Psychological Basis of Behavior	Pre-requisites: HUN100 Psychology as a science and its met...
HUL262	4	(3-1-0)	Environmental Psychology	Pre-requisites: HUN100 The nature and history of environme...

entry_id	
ee1060239	Abhishek Garg
ee1060311	Abhishek Minz
ee1060313	Amrit Singh Saini
ee1060316	Ankit Mittal
ee1060317	Anshul Khare
ee1060320	Deepesh Kumbhare
ee1060323	Gamandeep Sethi
ee1060325	Jatin Dev
ee1060326	Jatin Pasrija
ee1060328	Kinariwala Tejesh Rajesh
ee1060332	Milanshu Khandelwal

id	
sumantra	Sumantra Dutta Roy
sukumar	Sukumar Mishra
sumeet	Sumeet Agarwal
sprasad	Surinder Prasad
chandras	Suresh Chandra
skundu	Subiman Kundu
skak	Sunil Kak
svmodak	Subodh Vasant Modak
ghoshs	Sudarsan Ghosh
sudipto	Sudipto Mukherjee
suniljha	Sunil Jha

Methodology

- For community detection:
  - Topic Modeling Using LDA
  - Inferring User Interest.
  - Clustering
    - People grouping
    - Course grouping

- For course recommendation:
  - Collaborative Filtering using co-occurance matrix

TOPIC-WORD DISTRIBUTION:

Topic ID	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
0	physics	electron	plasma	quantum	magnetic	effects	thermal	devices	effect	semiconductor
1	linear	mal	theorem	functions	theory	integral	order	spaces	differential	function
2	process	production	growth	plant	micro	cell	microbial	kinetics	enzyme	formation
3	techniques	introduction	theory	application	advanced	review	examples	classical	theories	technique
4	optical	measurement	instruments	wave	instrumentation	measurements	optics	sensors	light	propagation
5	pre	requisites	ec	practical	training	industry	working	practices	end	credit
6	development	economic	international	policy	indian	global	india	role	public	growth
7	protein	structure	biology	biological	cell	proteins	function	membrane	molecular	introduction
8	control	basic	elements	features	electronic	controls	classification	controllers	feedback	hardware
9	topics	special	requisites	pre	research	ec	current	selected	development	areas
10	types	high	mechanisms	including	classification	bearings	friction	corrosion	forces	lubrication
11	computer	processes	tools	manufacturing	machine	assembly	metal	mel	forming	aided
12	laboratory	experiments	polymers	related	polymer	ii	size	preparation	polymerization	moulding
13	models	model	simulation	modeling	modelling	time	generation	level	simple	climate
14	chemical	chemistry	synthesis	reactions	methods	solid	organic	molecular	characterization	inorganic
15	science	social	human	hun	nature	theories	approaches	understanding	scientific	psychology
16	fabric	textile	principles	application	textiles	ttl	functional	structure	fabrics	properties
17	fibre	fibres	properties	structure	spinning	yarn	effect	parameters	temperature	process
18	applications	based	principles	laser	basics	technologies	fundamentals	lasers	display	description
19	quality	process	total	surface	introduction	geometric	line	standards	inspection	welding
20	noise	circuits	digital	signal	eel	modulation	frequency	basic	transmission	review
21	programming	algorithms	linear	networks	optimization	graph	problem	graphs	overlaps	algorithm
22	introduction	analysis	elements	mechanics	structural	vibration	stress	failure	strain	dynamic
23	power	system	operation	phase	distribution	load	performance	characteristics	line	transmission
24	machines	eel	electrical	requisites	electronics	motors	drives	applications	electric	pre
25	management	planning	maintenance	system	production	reliability	forecasting	approach	cycle	life

TOPIC AND WORD DISTRIBUTION RESULTS

Course	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
EEL702	0.011628	0.011628	0.023256	0.011628	0.011628	0.011628	0.011628	0.011628	0.023256	0.023256	0.023256
EEL703	0.006452	0.006452	0.006452	0.006452	0.006452	0.019355	0.012903	0.006452	0.03871	0.012903	0.012903
EEL704	0.010309	0.020619	0.010309	0.030928	0.010309	0.030928	0.010309	0.010309	0.103093	0.010309	0.010309
EEL705	0.007042	0.007042	0.007042	0.021127	0.007042	0.014085	0.014085	0.007042	0.098592	0.007042	0.007042
EEL706	0.009615	0.009615	0.009615	0.009615	0.009615	0.019231	0.009615	0.009615	0.019231	0.019231	0.028846
EEL707	0.013333	0.013333	0.013333	0.013333	0.026667	0.013333	0.013333	0.013333	0.026667	0.053333	0.013333
EEL708	0.012048	0.012048	0.012048	0.024096	0.012048	0.024096	0.012048	0.012048	0.012048	0.012048	0.012048
EEL709	0.009091	0.045455	0.009091	0.036364	0.009091	0.018182	0.009091	0.018182	0.009091	0.009091	0.009091
EEL710	0.012346	0.012346	0.012346	0.012346	0.012346	0.037037	0.012346	0.012346	0.024691	0.012346	0.012346
EEL711	0.007194	0.18705	0.035971	0.007194	0.007194	0.014388	0.007194	0.007194	0.007194	0.007194	0.007194
EEL712	0.006897	0.006897	0.006897	0.006897	0.137931	0.006897	0.013793	0.013793	0.02069	0.013793	0.013793

Entry_ID	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
eey127534	0.035777	0.07028	0.043841	0.06446	0.062563	0.06487	0.044706	0.043841	0.283837	0.035777	0.035777
eey127535	0.00813	0.01626	0.00813	0.01626	0.00813	0.03252	0.00813	0.00813	0.00813	0.04065	0.00813
eey127510	0.047449	0.055561	0.047985	0.062732	0.040076	0.040146	0.048317	0.039743	0.28344	0.031835	0.040076
eey127527	0.217696	0.079265	0.05295	0.094779	0.150866	0.068101	0.05295	0.078686	0.165955	0.234768	0.071301
eey127528	0.07348	0.14682	0.061743	0.107163	0.093603	0.170437	0.069034	0.070492	0.098411	0.192662	0.094833
eey127509	0.035553	0.078928	0.035553	0.103282	0.035553	0.059907	0.050994	0.10623	0.035553	0.061101	0.043273
eey127508	0.03446	0.114601	0.03446	0.091731	0.043894	0.07353	0.03446	0.043894	0.043894	0.20874	0.03446
eey127503	0.060848	0.111265	0.060848	0.112947	0.074954	0.153553	0.071133	0.107927	0.074954	0.316864	0.071388
eey127507	0.054657	0.123282	0.054657	0.117619	0.054657	0.171287	0.070336	0.126422	0.054657	0.255752	0.062496
eey127506	0.049746	0.09424	0.049746	0.123533	0.06308	0.103429	0.049746	0.058837	0.06308	0.140448	0.049746

STUDENT COMMUNITY MEMBERS FROM COURSE EEL709:

entry_id	name
mt5080454	Piyush Ahuja
mt5129056	Antoine Chassande Dit Patron
mt5100596	Himanshu Sharma
cs5090238	Ashish Gaurav
ee1090413	Sagar Sachdev
mt5080447	Gaurav Jaithlia
eey127509	Sumit Soman
eey127503	Arihant Jain
eey127507	Jobin Wilson
eey117511	Abhishek Srivastava
mt5060455	Vulluri Santhosh Kumar

entry_id	name
csy117547	Adhikaar Marwaha
ee1090384	Barun Kumar
ee5090491	Tanya Raghuvanshi
ee5090796	Kumar Sourav
ee5090478	Ashok Kumar Meghwal
ee5090476	Arashpreet Singh
ee5090477	Ashish Ranjan
ee5090475	Anurag Tiwari
ee5090472	Alok Singhal
ee5080420	Gaurav Soni
ee5090275	Akshay Gupta
cs1070172	Prasenjit
ee5090489	Shubham Gupta
ee5090488	Shaurya Pratap Singh
ee5090486	Rahul Gupta
ee5090484	Monalisa
ee5090482	Iftikhar Hussain Ansari
ee5090481	Honey Patodi
ee5090480	Gourav Naval
ee5080412	Alpana Singh
ee5090616	Sahil Garg
ee5090052	Aastha

entry_id	name
phz128488	Manish Anand
bsz128160	Pooja Gopal
bsz128161	Ravi Shankar
eey117503	Rama Kant Singh
ee5070407	Kamal Gupta
jvl112067	Jobin James
siy127526	Mr Sukhraj Singh
anz128203	Piyush Chanana
eez128123	Anukriti Bansal
eez128122	Sayak Bhattacharya
eez128124	Prabhmandeep Kaur
eez128128	Rakesh P
ee5080428	Swati Ganeti
jop112125	Manoj Gupta
eez118470	Shailesh Kumar
eey127508	Ramanjit Singh Ahuja
eey117513	Anubhuti Chopra
eez128378	Anupama Ray
eez128379	Shraddha Chaudhary
jop112619	Aniket Chowdhury

entry_id	
eet123000	Ben Mathew John
eet122878	Soniya Singhal
eet122984	Rahul Saluja
eet122981	Vishwanath Gaur
eet122980	Akash Kumar Agrawal
eet123059	Reeno Joseph Baby
eet122388	Narendra Vishwakarma
eet122387	Vaibhav Gaur
eet122072	Nitesh Gupta
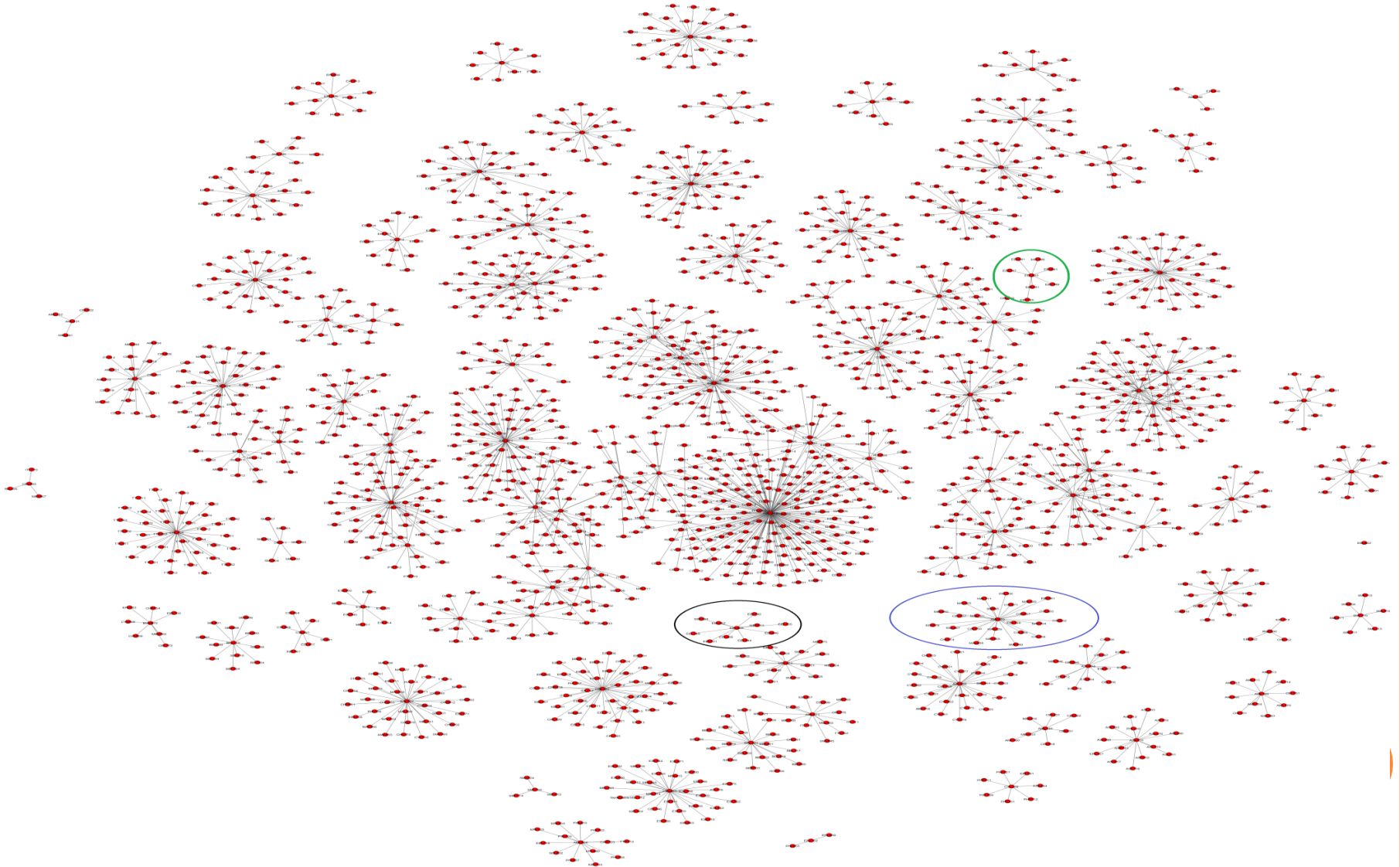
RECOMMENDATIONS:

course_id	title	credits	contents	name	name	entry_id
MAL701	Introduction to Programming and Data Structures	3	Definition of a programme. Programming methodology. Conc...	Niladri Chatterjee	Ben Mathew John	eet123000
EEP703	Computer Networks Lab	2	Simulation and Hardware Experiments on different aspects o...	Subrat Kar	Ben Mathew John	eet123000
EEP702	Software Laboratory	2	Students are expected to work under Windows and/or LINUX/...	Subrat Kar	Ben Mathew John	eet123000
EEP701	Digital System Lab	2	Students will design, implement and experiment with digital s...	Subrat Kar	Ben Mathew John	eet123000
EEL709	Pattern Recognition	3	Pre-requisites: MAL250 Introduction - What is Pattern Recog...	Sumeet Agarwal	Ben Mathew John	eet123000
EEL703	Computer Networks	3	Pre-requisites: EEL306 Overlaps with: CSL374, CSL858 Revi...	Subrat Kar	Ben Mathew John	eet123000
EEL702	Computer System Software	4	Pre-requisites: CSL201 & EEL308 Overlaps with: CSL728 Intr...	Santanu Chaudhury	Ben Mathew John	eet123000
EEL602	Operating Systems	4	Process management, Concurrent processes, mutual exclus...	Sumantra Dutta Roy	Ben Mathew John	eet123000
EEL601	Computer Architecture	3	High Performance Processor. Architecture. Motivation for Par...	Sumantra Dutta Roy	Ben Mathew John	eet123000
EED701	Minor Project (Computer Technology)	2		Sumantra Dutta Roy	Ben Mathew John	eet123000
CSL719	Synthesis of Digital Systems	4	Pre-requisites: CSL316 After a basic overview of the VLSI des...	Anshul Kumar	Ben Mathew John	eet123000

Recommendations	
Course ID	Course Name
EEL707	Multimedia Systems
EEL715	Image Processing
EEL711	Signal Theory
EEL731	Digital Signal Processing
EEC410	Colloquium
EEV704	Special Module in Computers
EEL319	Digital Signal Processing
EEL358	Operating Systems

RESULTS

70-30 Precision

70-30 Recall

70-30 f-measure
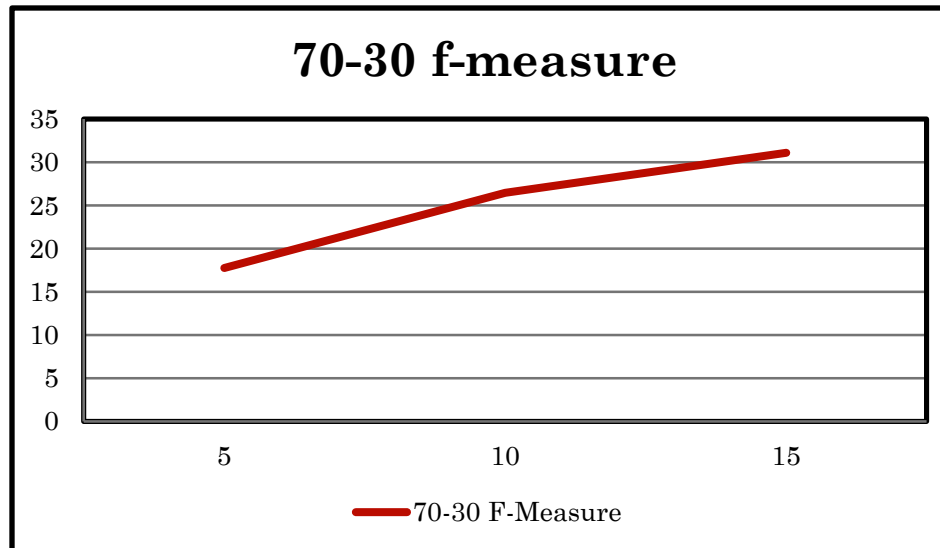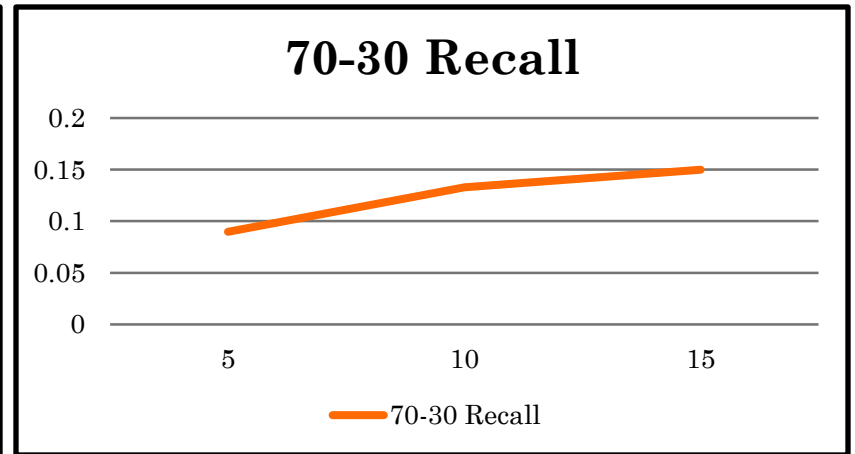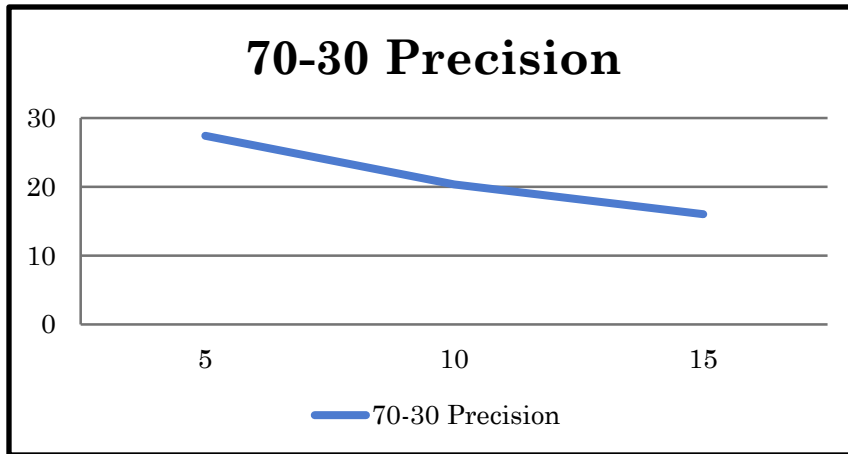
REFERENCES:

- Qi, Guo-Jun, Charu C. Aggarwal, and Thomas Huang. "Community Detection with Edge Content in Social Media Networks." *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. IEEE, 2012.

- Leskovec, Jure, Kevin J. Lang, and Michael Mahoney. "Empirical comparison of algorithms for network community detection." *Proceedings of the 19th international conference on World wide web*.ACM, 2010.

- McAuley, Julian, and Jure Leskovec. "Discovering social circles in ego networks." *arXiv preprint arXiv:1210.8182* (2012).

- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *the Journal of machine Learning research* 3 (2003): 993-1022.

- Owen, S., Anil, R., Dunning, T., & Friedman, E. (2011). *Mahout in action* (pp. 145-183). Manning.

- Mahout, Apache. "Scalable machine learning and data mining." (2012).

- Ingersoll, Grant. "Introducing apache mahou t." (2009).

- Negi, Sumit, and Santanu Chaudhury. "Predicting user-to-content links in Flickr Groups." *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*. IEEE, 2012.
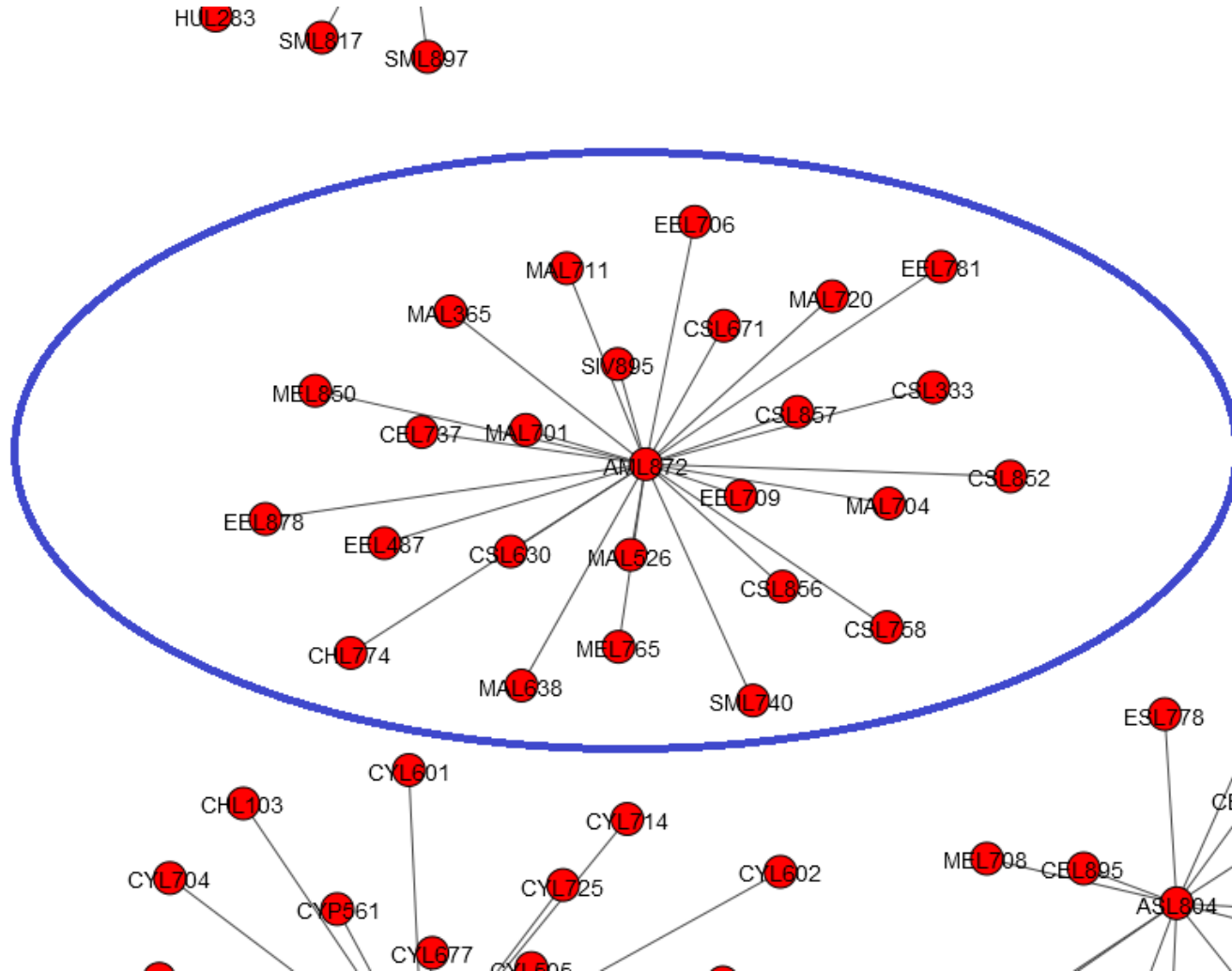
- http://www.ibm.com/developerworks/library/j-mahout/

- http://mallet.cs.umass.edu/topics.php

THANK YOU

QUESTIONS

APPENDIX 1:

RECOMMENDATION

course_id	title	credits	contents	name	name	entry_id
EEL709	Pattern Recognition	3	Pre-requisites: MAL250 Introduction - What is Pattern Recog...	Sumeet Agarwal	Geetanjali Srivastava	eez128304
EEL710	Coding Theory	3	Pre-requisites: EEL306 Measure of information; Source codin...	Ranjan Bose	Geetanjali Srivastava	eez128304
EEL774	Parameter Estimation and System Identification	3	Introduction and overview of Systems Identification, Adaptive ...	Rkp Bhatt	Geetanjali Srivastava	eez128304
EEL895	Broadband Communication and Information Systems	3	Fundamentals of telecom systems, Principles of communicat...	Vinod Chandra	Geetanjali Srivastava	eez128304
CRL704	Sensor Array Signal Processing	3	Representation of Space- Time Signals (Coordinate systems...	Arun Kumar	Geetanjali Srivastava	eez128304

Course id	Course name
EEL762	Digital Communication
EEL731	Digital Signal Processing 1
EEV711	Signal Theory
DTD899	Doctoral Thesis
EEV704	Robotics and Automation

RECOMMENDATION

course_id	title	credits	contents	name	name	entry_id
CSL105	Discrete Mathematical Structures	4	Overlaps with: MAL180 Fundamental structures: Functions (s...	Naveen Garg	Akshay Kumar	ee5080411
CSL356	Analysis and Design of Algorithms	4	Pre-requisites: CSL201 Overlaps with: MAL342 RAM model a...	Ragesh Jaiswal	Akshay Kumar	ee5080411
EEV704	Special Module in Computers	1	Pre-requisites: CSL201 and EC 90 Details will be decided by ...	Jayadeva	Akshay Kumar	ee5080411
EEL709	Pattern Recognition	3	Pre-requisites: MAL250 Introduction - What is Pattern Recog...	Sumeet Agarwal	Akshay Kumar	ee5080411
EED851	Major Project Part 1 (EI)	6	Pre-requisites: EC 165 This project spans also the course E...	Subrat Kar	Akshay Kumar	ee5080411
EED851	Major Project Part 1 (EI)	6	Pre-requisites: EC 165 This project spans also the course E...	S M K Rahman	Akshay Kumar	ee5080411
EED852	Major Project Part 2 (EI)	14	Pre-requisites: EED851 Continuation and completion of the ...	Kushal Kumar Shah	Akshay Kumar	ee5080411
MAL732	Financial Mathematics	4	Some Basic Definitions and Terminology; Basic option theory...	Suresh Chandra	Akshay Kumar	ee5080411

Course id	Course name
CSP301	Design practices in Computer Science
CSL302	Programming Languages
CSL211	Computer Architecture
ASL410	Numeric simulation of Atmospheric and Oceanic Phenomenon
MAL250	Introduction to Probability Theory and Stochastic Processes