CS5600 ASSIGNMENT 2

Citation Network Analysis

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Name	Roll Number
Deva Suvedh	CS22BTECH11016
Singa Divija Reddy	AI22BTECH11026
Medikonda Sreekar	CS22BTECH11037
Gujjala Vignesh	CS22BTECH11025
Nunavath Vishnu Teja	AI22BTECH11030

Contents

Overview

This report presents the analysis of a citation network derived from the DBLP-Citation-network V10 dataset. We perform link analysis to understand citation relationships, paper influence, and topical importance using PageRank-based methods.

Task 1: Graph Construction and Statistics

Dataset Filtering

We selected papers with at least 60 citations published between 2010 and 2015. Each paper is represented as a vertex, and a directed edge from paper X to paper Y indicates that X cites Y.

Graph Statistics

In this exercise, we construct a directed citation network where nodes represent papers and edges represent citations. Using NetworkX, we analyze structural graph properties such as connected components and the size of major subgraphs.

Metric	Value
Number of vertices	49572
Number of edges	163309
Number of weakly connected components (WCC)	6985
Number of strongly connected components (SCC)	47731
Nodes in largest WCC	41225
Edges in largest WCC	161635
Nodes in largest SCC	171
Edges in largest SCC	1190

Task 2: Paper Similarity Analysis

Co-citation and Bibliographic Coupling Scores

We compute two similarity measures between papers — Co-citation Score and Bibliographic Coupling Score — using the extracted citation graph. Below are the top-10 most similar paper pairs for each measure.

Observation

A high co-citation score indicates that two papers are often mentioned together by others, while a high bibliographic coupling score indicates that they share similar references, meaning they build upon similar prior work.

Co-citation Score:

Two papers A and B are said to be co-cited if they are cited together by other papers.

$$\text{CoCite}(A, B) = |\{P \mid P \text{ cites both } A \text{ and } B\}|$$

Top-10 Similar Papers based on Co-citation Score

S.No.	Title of Paper A	Title of Paper B
1	Real-time human pose recognition in parts from single depth images	Real-time human pose recognition in parts from single depth images
2	Very Deep Convolutional Networks for Large-Scale Image Recognition	ImageNet Classification with Deep Convolutional Neural Networks
3	DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition	ImageNet Classification with Deep Convolutional Neural Networks
4	Caffe: Convolutional Architecture for Fast Feature Embedding	ImageNet Classification with Deep Convolutional Neural Networks
5	OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks	ImageNet Classification with Deep Convolutional Neural Networks
6	Visualizing and Understanding Convolutional Networks	ImageNet Classification with Deep Convolutional Neural Networks
7	Very Deep Convolutional Networks for Large-Scale Image Recognition	Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation
8	Going deeper with convolutions	ImageNet Classification with Deep Convolutional Neural Networks
9	OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks	Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation
10	Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays	Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas

Bibliographic Coupling Score:

Two papers A and B are bibliographically coupled if they both cite the same papers.

$$\text{BiblioCouple}(A, B) = |\{P \mid A \text{ and } B \text{ both cite } P\}|$$

Top-10 Similar Papers based on Bibliographic Coupling Score

S.No.	Title of Paper A	Title of Paper B
1	Salient Object Detection: A Benchmark	Salient Object Detection: A Survey
2	Software-Defined Networking: A Comprehensive Survey	Security in Software Defined Networks: A Survey
3	Software-Defined Networking: A Comprehensive Survey	A Survey and a Layered Taxonomy of Software-Defined Networking
4	Design Guidelines for Spatial Modulation	Spatial Modulation for Generalized MIMO: Challenges, Opportunities, and Implementation
5	Software-Defined Networking: A Comprehensive Survey	A Survey on Software-Defined Networking
6	Urban Computing: Concepts, Methodologies, and Applications	Trajectory Data Mining: An Overview
7	Salient Object Detection: A Survey	Salient Object Detection: A Discriminative Regional Feature Integration Approach
8	A survey on information visualization: recent advances and challenges	A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges
9	Deep learning of representations: looking forward	Representation Learning: A Review and New Perspectives
10	Salient Object Detection: A Benchmark	Salient Object Detection: A Discriminative Regional Feature Integration Approach

Task 3: PageRank Damping Factor Evaluation

Graph Statistics

Metric	Value
Nodes	49572
Edges	163309

Pearson Correlation vs Damping Factor (d)

Damping Factor (d)	Pearson Correlation
0.15	0.785755
0.25	0.768759
0.35	0.745674
0.45	0.732333
0.55	0.712790
0.65	0.688880
0.75	0.674215
0.85	0.554154
0.95	0.640048

Table 1: Correlation Summary for Different Damping Factors

Top-10 Papers ($d = 0.15$)

S.No.	Title	PageRank Score
1	LIBSVM: A library for support vector machines	0.000848
2	The Pascal Visual Object Classes (VOC) Challenge	0.000275
3	Object Detection with Discriminatively Trained Part-Based Models	0.000223
4	Community detection in graphs	0.000160
5	Fast and Scalable Local Kernel Machines	0.000146
6	What is Twitter, a social network or a news media?	0.000143
7	Reducibility Among Combinatorial Problems	0.000131
8	Talking about tactile experiences	0.000110
9	ImageNet Classification with Deep Convolutional Neural Networks	0.000103
10	KEGG for representation and analysis of molecular networks involving d	0.000100

Table 2: Top-10 Papers for $d = 0.15$

Top-10 Papers ($d = 0.85$)

S.No.	Title	PageRank Score
1	LIBSVM: A library for support vector machines	0.008925
2	Fast and Scalable Local Kernel Machines	0.007273
3	The Pascal Visual Object Classes (VOC) Challenge	0.004482
4	Factored Shapes and Appearances for Parts-based Object Understanding	0.003910
5	Object Detection with Discriminatively Trained Part-Based Models	0.002628
6	ClassCut for unsupervised class segmentation	0.002193
7	Community detection in graphs	0.001018
8	A Singular Value Thresholding Algorithm for Matrix Completion	0.000959
9	TwitterRank: finding topic-sensitive influential twitterers	0.000948
10	Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nucle	0.000903

Table 3: Top-10 Papers for $d = 0.85$

Summary

Best d	0.15 (Correlation = 0.785755)
Worst d	0.85 (Correlation = 0.554154)

Table 4: Best and Worst Damping Factors

Task 4: Topic-Sensitive PageRank Results

Theoretical Comparison

- **Normal PageRank:** Uses a uniform teleportation probability. The random surfer model assumes equal chance of jumping to any node.

$$PR = dA'PR + (1 - d)\frac{1}{N}$$

- **Topic-Sensitive PageRank:** Teleportation is biased towards nodes related to a given topic. A separate PageRank vector is computed for each topic.

- **Personalized PageRank:** Similar to topic-sensitive but the personalization vector is based on a specific user or preference set. Commonly used in recommendation systems.

Implementation and Results

The topic-sensitive PageRank algorithm was implemented with damping factor $d = 0.85$ for the following topics: *security, hashing, streaming, timeseries, search*. A paper is considered relevant if its title contains the topic keyword.

Graph Statistics

Metric	Value
Nodes	49572
Edges	163309

Top-10 Papers for Topic: Security

S.No.	Title	PageRank Score	Citations
1	Security and Privacy Challenges in Cloud Computing Environments	8.956396e-03	33
2	Neutralization: new insights into the problem of employee systems secu	8.016105e-03	21
3	SecureCloud: Towards a Comprehensive Security Framework for Cloud Comp	7.317813e-03	3
4	Enabling Public Auditability and Data Dynamics for Storage Security in	6.750305e-03	47
5	Dependable and Secure Sensor Data Storage with Dynamic Integrity Assur	6.745501e-03	3
6	Stackelberg vs. Nash in security games: interchangeability, equivalenc	6.059331e-03	8
7	A lattice-based approach to mashup security	5.658768e-03	8
8	Google Android: A Comprehensive Security Assessment	5.487063e-03	22
9	Permissive dynamic information flow analysis	5.161266e-03	6
10	Improving Wireless Physical Layer Security via Cooperating Relays	5.131490e-03	46

Table 5: Top-10 Papers for Topic: Security

Top-10 Papers for Topic: Hashing

S.No.	Title	PageRank Score	Citations
1	Semi-supervised hashing for scalable image retrieval	4.883837e-02	41
2	Sequential Projection Learning for Hashing with Compact Codes	3.323561e-02	31
3	SPEC hashing: Similarity preserving algorithm for entropy-based coding	2.229536e-02	14
4	Hashing with Graphs	1.938876e-02	39
5	Weakly-supervised hashing in kernel space	1.879445e-02	16
6	Minimal Loss Hashing for Compact Binary Codes	1.811736e-02	42
7	Hashing Algorithms for Large-Scale Learning	1.810516e-02	4
8	Self-taught hashing for fast similarity search	1.756456e-02	24
9	Supervised hashing with kernels	1.665899e-02	39
10	b-Bit minwise hashing	1.511820e-02	7

Table 6: Top-10 Papers for Topic: Hashing

Top-10 Papers for Topic: Streaming

S.No.	Title	PageRank Score	Citations
1	An evaluation of TCP-based rate-control algorithms for adaptive intern	2.727341e-02	9
2	An experimental evaluation of rate-adaptation algorithms in adaptive s	1.712650e-02	31
3	An experimental investigation of the Akamai adaptive video streaming	1.420783e-02	10
4	Watching Video over the Web: Part 1: Streaming Protocols	9.584491e-03	6
5	Rate adaptation for adaptive HTTP streaming	8.557227e-03	12
6	On the exact space complexity of sketching and streaming small norms	7.169067e-03	4
7	Feedback control for adaptive live video streaming	6.902230e-03	12
8	Impact of Network Dynamics on User's Video Quality: Analytical Framewo	6.879428e-03	3
9	UUSee: Large-Scale Operational On-Demand Streaming with Random Network	6.624422e-03	8
10	The MPEG-DASH Standard for Multimedia Streaming Over the Internet	6.561158e-03	8

Table 7: Top-10 Papers for Topic: Streaming

Top-10 Papers for Topic: Timeseries

S.No.	Title	PageRank Score	Citations
1	An artificial neural network (p,d,q) model for timeseries forecasting	9.629833e-01	3
2	LIBSVM: A library for support vector machines	3.702633e-03	638
3	Fast and Scalable Local Kernel Machines	2.387390e-03	3
4	Factored Shapes and Appearances for Parts-based Object Understanding	2.106180e-03	19
5	ClassCut for unsupervised class segmentation	2.019267e-03	9
6	The Pascal Visual Object Classes (VOC) Challenge	1.121098e-03	376
7	Object Detection with Discriminatively Trained Part-Based Models	9.311353e-04	430
8	Random Walks, Markov Processes and the Multiscale Modular Organization	3.297217e-04	11
9	A Singular Value Thresholding Algorithm for Matrix Completion	2.996145e-04	134
10	Fixed point and Bregman iterative methods for matrix rank minimization	2.804471e-04	59

Table 8: Top-10 Papers for Topic: Timeseries

Top-10 Papers for Topic: Search

S.No.	Title	PageRank Score	Citations
1	LIBSVM: A library for support vector machines	9.622692e-03	638
2	Fast and Scalable Local Kernel Machines	6.252425e-03	3
3	Factored Shapes and Appearances for Parts-based Object Understanding	5.317816e-03	19
4	The Pascal Visual Object Classes (VOC) Challenge	5.055265e-03	376
5	ClassCut for unsupervised class segmentation	3.664222e-03	9
6	Object Detection with Discriminatively Trained Part-Based Models	2.847288e-03	430
7	Beyond DCG: user behavior as a predictor of a successful search	2.759189e-03	11
8	A Theoretical and Empirical Study of Search-Based Testing: Local, Glob	2.393072e-03	20
9	Action design research	2.359597e-03	9
10	A Dynamic Model of Sponsored Search Advertising	2.326893e-03	10

Table 9: Top-10 Papers for Topic: Search