Machine Learning Assignment 3

Sayan Sinha
16CS10048

INTRODUCTION

Unsupervised machine learning generally boils down to grouping similar objects based on some common properties shared among them. One of the most popular methods of doing so is the k-means clustering algorithm. One common problem with the k-means is that the clusters are globular, hence cannot be used to model arbitrary-shaped clusters. In this assignment however, we explore the methods of hierarchical and graph-based clustering algorithms.

MODEL

Hierarchical clustering

The assignment involves finding out clusters by initially considering each node to be a single cluster and merging them one by one till we get the optimal number of clusters desired (bottom-up) [2]. Here, every individual paper is considered to be a node. The clusters are merged based on their proximities. In every iteration the closest two clusters are merged. The proximity between two nodes is computed using the Jaccard coefficient [1]:

$$JC_{HS} = JC_{SH} = \frac{|H \cap S|}{|H \cup S|}$$

where H and S are two distinct nodes. The comparisons are made among the "Topics" of the papers.

Proximity between two clusters is taken as the minimum Jaccard coefficient among the nodes of each cluster in the case of complete linkage, while the maximum Jaccard coefficient is used for single linkage. In this experiment we use a separate function to compute the proximity. While merging two nodes, we delete the existing nodes and append the newly created node to the end in some cases. In other cases, we replace one of the existing clusters with the merged cluster and delete the other one. The placement of the newly created node has created various qualities of the clusters which have been discussed later.

Graph based clustering

In the assignment, we used a specific graph based method known as the Girvan-Newman clustering algorithm [3]. Here, we consider each paper to be a node and join an edge between the ones having a Jaccard coefficient greater than a certain coefficient. The connected components of the graph gives us the clusters. In every iteration, we find the edge with the highest centrality and remove it. We proceed in this way till we get the desired number of clusters (or connected components).

EXPERIMENTATION

Part 1

The AAAI dataset is provided in which we are to cluster on the basis of the "Topics". There are nine unique "High level topics", only one of which has been assigned to each paper, as per the dataset. Therefore we try to find out nine clusters and then aim to find out how good the clustering was compared to the given high level domain. We compute this

measure of goodness using the NMI score for a set of clusters. NMI score is given as:

$$NMI(Y,C) = \frac{2 \times I(Y;C)}{[H(Y) + H(C)]}$$

where,

Y = class labels, C = cluster labels, $H(.)$ = Entropy and $I(Y;C)$ = Mutual Information between Y and C, where

$$I(Y;C) = H(Y) - H(Y|C)$$

where $H(Y|C)$ is the entropy of class Y in cluster C.

In the case of complete linkage, the size of the clusters obtained were:
[11, 11, 14, 12, 16, 18, 16, 23, 29].

The NMI value of the cluster was: **0.387**.

For single linkage, they were:
[1, 1, 1, 1, 1, 42, 47, 55, 1].

The NMI value of the cluster was: **0.509**.

One important observation to note here would be that the placement of the newly created node created a huge difference in the NMI values obtained. If we place the newly generated cluster at the end of the list of nodes and delete the existing ones, we get the NMI values as 0.387 (complete) and 0.509 (single). Instead of placing the newly generated cluster at the end of the list of clusters, if we would instead replace the first cluster with the merged cluster and delete the second, we get the NMI values of 0.211 (complete) and 0.107 (single). If instead we replace the second cluster with the merged one and delete the first, we get 0.236 (complete) and 0.509 (single). Hence, using validation, we find that placing at the back of the list works best for complete and replacing the second cluster works best for single, and we choose such values accordingly.

Part 2

The python package *networkx* was used to create and perform operations on the graph. The centralities were also found out using an inbuilt function. Otherwise, centrality is given by:

$$BC(e) = \sum_{v,w \in V} \frac{b_{vw}(e)}{b_{vw}}$$

where $b_{vw}(e)$ is the number of shortest paths between the nodes v and w through the edge e. b_{vw} is given by the total number of shortest paths between v and w.

Varying the threshold of Jaccard coefficient for initial linkage by 0.01 from 0 to 0.2, we obtain a curve as shown in Fig 1. Thus, we see the best threshold is 0.15 which gives an NMI score of **0.627**.

The size of the clusters obtained using the graph based method, setting the threshold to 0.15 is:
[40, 39, 22, 19, 7, 11, 1, 10, 1].

The cluster map has been presented in Fig 2.

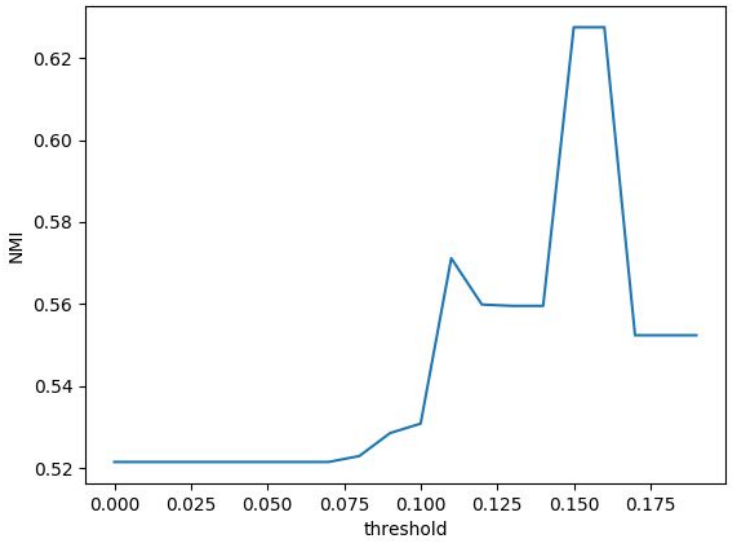

Fig 1: NMI vs threshold

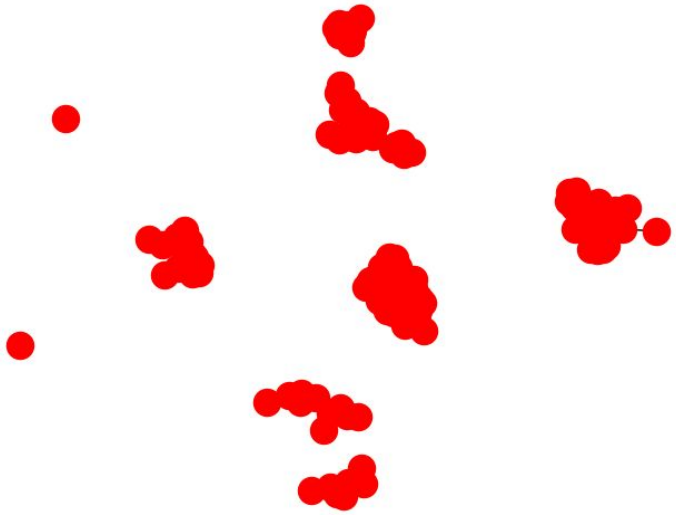

Fig 2: Map of clusters having the maximum determined NMI

CLUSTERS

Complete linkage

Cluster no 1:

['A Topic-Based Coherence Model for Statistical Machine Translation', 'An Extended GHKM Algorithm for Inducing Lambda-SCFG', 'Filtering with Logic Programs and its Application to General Game Playing', 'Dynamic Minimization of Sentential Decision Diagrams', 'Effective Bilingual Constraints for Semi-supervised Learning of Named Entity Recognizers', 'Joint inference of extraction and labelling via graph propagation for dictionary construction', 'Probabilistic Sense Sentiment Similarity through Hidden Emotions', 'Guiding Scientific Discovery with Explanations', 'Towards Cohesive Anomalies Mining', 'From Interest to Function: Location Estimation in Social Media', 'Time-dependent Trajectory Regression on Road Networks via Multi-Task Learning']

Cluster no 2:
['Video Saliency Detection via Dynamic Consistent Spatio-Temporal Attention Modelling', 'Teaching Classification Boundaries to Humans', 'Sparse Multi-task Learning for Detecting Influential Nodes in an Implicit Diffusion Network', 'Scalable Lifelong Learning with Active Task Selection', 'Multi-agent Knowledge and Belief Change in the Situation Calculus', 'Causal Transportability with Limited Experiments', 'Progression of Decomposed Situation Calculus Theories', 'm-Transportability: Transportability of Causal Effects from Multiple Environments', 'Ranking Scientific Articles by Exploiting Citations, Authors, Journals and Time Information', 'Unified Constraint Propagation on Multi-View Data', 'Walking on Minimax Paths for k-NN Search']

Cluster no 3:
['Uncorrelated Lasso', 'SMILe: Shuffled Multiple-Instance Learning', 'A Maximum K-Min Approach for Classification', 'Projected Group Sparse Coding for Image Classification', 'Analyzing the effectiveness of adversary modeling in security games', 'An Agent Design for Repeated Negotiation and Information Revelation with People', 'Social Rankings in Human-Computer Committees', 'Continuous Conditional Random Fields for Efficient Regression in Large Fully Connected Graphs', 'Large-Scale Hierarchical Classification via Stochastic Perceptron', 'HC-Search: Learning Heuristics and Cost Functions for Structured Prediction', 'Search More, Disclose Less', 'Multi-Cycle Query Caching in Agent Programming', 'Interdependent Multi-Issue Negotiation for Energy Exchange in Remote Communities', 'Information Sharing Under Costly Communication in Joint Exploration']

Cluster no 4:
['Reciprocal Hash Tables for Nearest Neighbor Search', 'Robust Discrete Matrix Completion', 'Automated Workflow Synthesis', 'Sensitivity of diffusion dynamics to network uncertainty', 'Multi-Armed Bandit with Budget Constraint and Variable Costs', 'The cascade auction – a mechanism for deterring collusion in auctions', 'Posted Prices Exchange for Display Advertising Contracts', 'On the Value of using Group Discounts under Price Competition', 'A Robust Bayesian Truth Serum for Non-binary Signals', 'Simple Temporal Problems with Taboo Regions', 'Improving the Performance of Consistency Algorithms by Localizing and Bolstering Propagation in a Tree Decomposition', 'A Morphogenetically Assisted Design Variation Tool']

Cluster no 5:
['A Kernel Density Estimate-based approach to Component Goodness Modeling', "Reasoning about Conditional Independence under Uncertainty: Axioms, Algorithms and Levesque's Situations to the Rescue", 'A Fast Pairwise Heuristic for Planning under Uncertainty', 'Complexity of Inferences in Polytree-shaped Semi-Qualitative Probabilistic Networks', 'GiSS: Combining Gibbs Sampling and SampleSearch for Inference in Mixed Probabilistic and Deterministic Graphical Models', 'Boosting Lifted Likelihood Maximization for MAP Inference by Virtual Evidence', 'Integrating Programing by Example and Natural Language Programing', 'Grounding Natural Language References to Unvisited and Hypothetical Location', 'The Automated Acquisition of Suggestions from Tweets', 'A Pattern Matching Based Graphical Model for Question Subjectivity Prediction', 'Generating Natural-Language Video Descriptions Using Text-Mined Knowledge', 'A Hierarchical Aspect-Sentiment Model for Online Reviews', 'Automatic Identification of Conceptual Metaphors With Limited Knowledge', 'Symmetry-Aware Marginal Density Estimation', 'Supervised Nonnegative Tensor Factorization with Maximum-Margin Constraints', 'Supervised Coupled Dictionary Learning with Group Structures for Multi-modal Retrieval']

Cluster no 6:
['Cost-Optimal Planning by Self-Interested Agents', 'On the Social Welfare of Mechanisms for Repeated Batch Matching', 'How to Cut a Cake Before the Party Ends', 'Bundling Attacks in Judgment Aggregation', 'Learning Integrated Symbolic and Continuous Action Models for Continuous Domains', 'Spectral Rotation vs K-means in Spectral Clustering', 'Clustering with Complex Constraints - Algorithms and Applications', 'Discovering hierarchical structure for sources and entities', 'Basis Adaptation for Sparse Nonlinear Reinforcement Learning', 'Structured Kernel-Based Reinforcement Learning', 'Sample Complexity and Performance Bounds for Non-parametric Approximate Linear Programming', 'PAC Optimal Exploration in Continuous Space Markov Decision Processes', 'Pruning for Monte Carlo Distributed Reinforcement Learning in Decentralized POMDPs', 'Modelling and Control of Mixed Observability Multiagent Systems', 'Vector-valued Multi-view Semi-supervised Learning for Multi-label Image Classification', 'A Tensor-Variate Gaussian Process for Classification of Multidimensional Structured Data', 'Lazy Gaussian Process Committee for Real-Time Online Regression', 'On Power law kernels, corresponding Reproducing Kernel Hilbert Space and applications']

Cluster no 7:
['Salient Object Detection via Low-Rank and Structured Sparse Matrix Decomposition', 'Joint Object and Pose Recognition using Homeomorphic Manifold Analysis', 'Vesselness Features and the Inverse Compositional AAM for Robust Face Recognition using Thermal IR', 'Gradient Networks for Shape-Based Object Instance Detection', 'Incremental Learning Framework for Indoor Scene Recognition', 'Abstract Preference Frameworks — a Unifying Perspective on Separability and Strong Equivalence', 'Radial Restraint: A Semantically Clean Approach to Bounded Rationality for Logic Programs', 'A General Formal Framework for Pathfinding Problems with Multiple Agents using Answer Set Programming', 'Liberal Safety for Answer Set Programs with External Sources', 'Backdoors to Normality for Disjunctive Logic Programs', 'Enforcing Meter in Finite-Length Markov Sequences', 'Temporal Milestones in HTNs', 'When is Brute-Force Avoidable for CSP?', 'Extending STR to a Higher-Order Consistency', 'Answering Counting Aggregate Queries over Ontologies of DL-Lite Family', 'Story Generation with Crowdsourced Plot Graphs']

Cluster no 8:
['A Generalized Student-t Based Approach to Mixed-Type Anomaly Detection', 'Unsupervised Cluster Matching via Probabilistic Latent Variable Models', 'Convex Subspace Representation Learning from Multi-view Data', 'Deep Manifold Learning', 'A Cyclic Weighted Median Method for L1 Low-Rank Matrix Factorization with Missing Entries', 'Parameterized Complexity Results for Case-Based Planning', 'Qualitative Planning under Partial Observability in Multi-Agent Domains', 'A First-Order Formalization of Commitments and Goals', 'Data-Parallel Computing Meets STRIPS', 'Model-Lite Case-Based Planning', 'Assumption-Based Planning: Generating plans and explanations under incomplete knowledge', 'Hypothesis Exploration for Malware Detection using Planning', 'Mixed Heuristic Local Search for Protein Structure Prediction', 'External Memory Best-First Search for Multiple Sequence Alignment', 'A Robust Bidirectional Search Using Heuristic Improvement', 'Red-Black Relaxed Plan Heuristics', 'Truncated LPA* : Faster Replanning by Exploiting Suboptimality', 'RockIt: Exploiting Parallelism and Symmetry for MAP Inference in Statistical Relational Models', 'Greedy or Not? Local search move selection for MAXSAT', 'Partial MUS Enumeration', 'Resolution and Parallelizability: Barriers to the Efficient Parallelization of SAT

Solvers', 'Improving WalkSAT for Random k-Satisfiability Problem with $k>3$', 'Domain-specific Heuristics in Answer Set Programming']

Cluster no 9:
['Optimizing Objective Function Parameters for Strength in Computer Game-Playing', 'Improved Optimal Search Heuristics with Manifold Learning', 'A Concave Conjugate Approach for Nonconvex Penalized Regression with the MCP Penalty', 'Online Lazy Updates for Portfolio Selection with Transaction Costs', 'Formalizing Hierarchical Clustering as Integer Linear Programming', 'Smart Multi-task Bregman Clustering and Multi-task Kernel Clustering', 'Instructor Rating Markets', 'Optimal Coalition Structures in Cooperative Graph Games', 'Composition Games for Distributed Systems: the EU Grant games', 'Bounding the Cost of Stability in Games over Interaction Networks', 'Strategic Behavior when Allocating Indivisible Goods Sequentially', 'Efficient evolutionary dynamics with extensive-form games', 'Automating Collusion Detection in Sequential Games', 'How Bad is Selfish Voting?', 'Equilibria of Online Scheduling Algorithms', 'Fast Equilibrium Computation for Infinitely Repeated Games', 'Algorithms for strong Nash equilibrium with more than two agents', 'Solving security games on graphs via marginal probabilities', 'Rank Aggregation via Low-rank and Structured-sparse Decomposition', 'Multi-Label Learning with PRO Loss', 'Teamwork with Limited Knowledge of Teammates', 'Multiagent Learning with a Noisy Global Reward Signal', 'Bribery in Voting With Soft Constraints', 'A Framework for Aggregating Influenced CP-nets and its Resistance to Bribery', 'Ties Matter: Complexity of Manipulation when Tie-breaking with a Random Vote', 'Computational Aspects of Nearly Single-Peaked Electorates', 'Dynamic Social Choice: Foundations and Algorithms', 'Timelines with Uncontrollability', 'Decoupling the Multiagent Disjunctive Temporal Problem']

Single Linkage

Cluster no 1:
['Cost-Optimal Planning by Self-Interested Agents']

Cluster no 2:
['Enforcing Meter in Finite-Length Markov Sequences']

Cluster no 3:
['Answering Counting Aggregate Queries over Ontologies of DL-Lite Family']

Cluster no 4:
['A Kernel Density Estimate-based approach to Component Goodness Modeling']

Cluster no 5:
['Learning Integrated Symbolic and Continuous Action Models for Continuous Domains']

Cluster no 6:
['Model-Lite Case-Based Planning', 'Data-Parallel Computing Meets STRIPS', 'Parameterized Complexity Results for Case-Based Planning', 'Hypothesis Exploration for Malware Detection using Planning', 'Assumption-Based Planning: Generating plans and explanations under incomplete knowledge', 'Online Lazy Updates for Portfolio Selection with Transaction Costs', 'Resolution and

Parallelizability: Barriers to the Efficient Parallelization of SAT Solvers', 'Partial MUS Enumeration', 'Greedy or Not? Local search move selection for MAXSAT', 'Truncated LPA* : Faster Replanning by Exploiting Suboptimality', 'Complexity of Inferences in Polytree-shaped Semi-Qualitative Probabilistic Networks', 'Boosting Lifted Likelihood Maximization for MAP Inference by Virtual Evidence', 'GiSS: Combining Gibbs Sampling and SampleSearch for Inference in Mixed Probabilistic and Deterministic Graphical Models', 'RockIt: Exploiting Parallelism and Symmetry for MAP Inference in Statistical Relational Models', 'Symmetry-Aware Marginal Density Estimation', 'A Concave Conjugate Approach for Nonconvex Penalized Regression with the MCP Penalty', 'Improved Optimal Search Heuristics with Manifold Learning', 'Optimizing Objective Function Parameters for Strength in Computer Game-Playing', 'Mixed Heuristic Local Search for Protein Structure Prediction', 'A Robust Bidirectional Search Using Heuristic Improvement', 'External Memory Best-First Search for Multiple Sequence Alignment', 'Dynamic Minimization of Sentential Decision Diagrams', 'Filtering with Logic Programs and its Application to General Game Playing', 'Multi-agent Knowledge and Belief Change in the Situation Calculus', 'm-Transportability: Transportability of Causal Effects from Multiple Environments', 'Progression of Decomposed Situation Calculus Theories', 'Causal Transportability with Limited Experiments', 'Extending STR to a Higher-Order Consistency', 'When is Brute-Force Avoidable for CSP?', 'A Morphogenetically Assisted Design Variation Tool', 'Improving the Performance of Consistency Algorithms by Localizing and Bolstering Propagation in a Tree Decomposition', 'Domain-specific Heuristics in Answer Set Programming', 'Improving WalkSAT for Random k-Satisfiability Problem with $k>3$', 'Temporal Milestones in HTNs', 'Simple Temporal Problems with Taboo Regions', 'Red-Black Relaxed Plan Heuristics', 'A Fast Pairwise Heuristic for Planning under Uncertainty', "Reasoning about Conditional Independence under Uncertainty: Axioms, Algorithms and Levesque's Situations to the Rescue", 'A First-Order Formalization of Commitments and Goals', 'Qualitative Planning under Partial Observability in Multi-Agent Domains', 'Decoupling the Multiagent Disjunctive Temporal Problem', 'Timelines with Uncontrollability']

Cluster no 7:
['Strategic Behavior when Allocating Indivisible Goods Sequentially', 'How Bad is Selfish Voting?', 'Bundling Attacks in Judgment Aggregation', 'How to Cut a Cake Before the Party Ends', 'Ties Matter: Complexity of Manipulation when Tie-breaking with a Random Vote', 'Dynamic Social Choice: Foundations and Algorithms', 'Computational Aspects of Nearly Single-Peaked Electorates', 'A Framework for Aggregating Influenced CP-nets and its Resistance to Bribery', 'Bribery in Voting With Soft Constraints', 'Solving security games on graphs via marginal probabilities', 'Algorithms for strong Nash equilibrium with more than two agents', 'A Robust Bayesian Truth Serum for Non-binary Signals', 'On the Value of using Group Discounts under Price Competition', 'Posted Prices Exchange for Display Advertising Contracts', 'The cascade auction – a mechanism for deterring collusion in auctions', 'Composition Games for Distributed Systems: the EU Grant games', 'Optimal Coalition Structures in Cooperative Graph Games', 'Bounding the Cost of Stability in Games over Interaction Networks', 'Fast Equilibrium Computation for Infinitely Repeated Games', 'Equilibria of Online Scheduling Algorithms', 'Automating Collusion Detection in Sequential Games', 'Efficient evolutionary dynamics with extensive-form games', 'Multi-Cycle Query Caching in Agent Programming', 'Search More, Disclose Less', 'Information Sharing Under Costly Communication in Joint Exploration', 'Interdependent Multi-Issue Negotiation for Energy Exchange in Remote Communities', 'Social Rankings in Human-Computer Committees', 'An Agent Design for Repeated Negotiation and Information Revelation with People', 'PAC Optimal Exploration in Continuous Space Markov Decision Processes', 'Sample Complexity and Performance Bounds for Non-parametric

Approximate Linear Programming', 'Structured Kernel-Based Reinforcement Learning', 'Basis Adaptation for Sparse Nonlinear Reinforcement Learning', 'Modelling and Control of Mixed Observability Multiagent Systems', 'Pruning for Monte Carlo Distributed Reinforcement Learning in Decentralized POMDPs', 'Multiagent Learning with a Noisy Global Reward Signal', 'Teamwork with Limited Knowledge of Teammates', 'On the Social Welfare of Mechanisms for Repeated Batch Matching', 'Instructor Rating Markets', 'Automated Workflow Synthesis', 'Multi-Armed Bandit with Budget Constraint and Variable Costs', 'Analyzing the effectiveness of adversary modeling in security games', 'A General Formal Framework for Pathfinding Problems with Multiple Agents using Answer Set Programming', 'Backdoors to Normality for Disjunctive Logic Programs', 'Liberal Safety for Answer Set Programs with External Sources', 'Radial Restraint: A Semantically Clean Approach to Bounded Rationality for Logic Programs', 'Abstract Preference Frameworks — a Unifying Perspective on Separability and Strong Equivalence', 'Sensitivity of diffusion dynamics to network uncertainty']

Cluster no 8:
['Grounding Natural Language References to Unvisited and Hypothetical Location', 'A Pattern Matching Based Graphical Model for Question Subjectivity Prediction', 'The Automated Acquisition of Suggestions from Tweets', 'Automatic Identification of Conceptual Metaphors With Limited Knowledge', 'A Hierarchical Aspect-Sentiment Model for Online Reviews', 'Probabilistic Sense Sentiment Similarity through Hidden Emotions', 'Joint inference of extraction and labelling via graph propagation for dictionary construction', 'Effective Bilingual Constraints for Semi-supervised Learning of Named Entity Recognizers', 'Unsupervised Cluster Matching via Probabilistic Latent Variable Models', 'A Generalized Student-t Based Approach to Mixed-Type Anomaly Detection', 'Integrating Programing by Example and Natural Language Programing', 'An Extended GHKM Algorithm for Inducing Lambda-SCFG', 'A Topic-Based Coherence Model for Statistical Machine Translation', 'Convex Subspace Representation Learning from Multi-view Data', 'A Cyclic Weighted Median Method for L1 Low-Rank Matrix Factorization with Missing Entries', 'Deep Manifold Learning', 'Sparse Multi-task Learning for Detecting Influential Nodes in an Implicit Diffusion Network', 'Generating Natural-Language Video Descriptions Using Text-Mined Knowledge', 'Joint Object and Pose Recognition using Homeomorphic Manifold Analysis', 'Salient Object Detection via Low-Rank and Structured Sparse Matrix Decomposition', 'Video Saliency Detection via Dynamic Consistent Spatio-Temporal Attention Modelling', 'Incremental Learning Framework for Indoor Scene Recognition', 'Gradient Networks for Shape-Based Object Instance Detection', 'Vesselness Features and the Inverse Compositional AAM for Robust Face Recognition using Thermal IR', 'Continuous Conditional Random Fields for Efficient Regression in Large Fully Connected Graphs', 'Scalable Lifelong Learning with Active Task Selection', 'Discovering hierarchical structure for sources and entities', 'Clustering with Complex Constraints - Algorithms and Applications', 'Time-dependent Trajectory Regression on Road Networks via Multi-Task Learning', 'Towards Cohesive Anomalies Mining', 'Guiding Scientific Discovery with Explanations', 'From Interest to Function: Location Estimation in Social Media', 'Spectral Rotation vs K-means in Spectral Clustering', 'Uncorrelated Lasso', 'A Tensor-Variate Gaussian Process for Classification of Multidimensional Structured Data', 'Supervised Coupled Dictionary Learning with Group Structures for Multi-modal Retrieval', 'Supervised Nonnegative Tensor Factorization with Maximum-Margin Constraints', 'Vector-valued Multi-view Semi-supervised Learning for Multi-label Image Classification', 'Teaching Classification Boundaries to Humans', 'On Power law kernels, corresponding Reproducing Kernel Hilbert Space and applications', 'Lazy Gaussian Process Committee for Real-Time Online Regression', 'Smart Multi-task Bregman Clustering and Multi-task Kernel Clustering', 'Walking on Minimax Paths for k-NN Search',

'Unified Constraint Propagation on Multi-View Data', 'HC-Search: Learning Heuristics and Cost Functions for Structured Prediction', 'Large-Scale Hierarchical Classification via Stochastic Perceptron', 'SMILe: Shuffled Multiple-Instance Learning', 'Multi-Label Learning with PRO Loss', 'Rank Aggregation via Low-rank and Structured-sparse Decomposition', 'Projected Group Sparse Coding for Image Classification', 'A Maximum K-Min Approach for Classification', 'Ranking Scientific Articles by Exploiting Citations, Authors, Journals and Time Information', 'Formalizing Hierarchical Clustering as Integer Linear Programming', 'Reciprocal Hash Tables for Nearest Neighbor Search', 'Robust Discrete Matrix Completion']

Cluster no 9:
['Story Generation with Crowdsourced Plot Graphs']

Graph based clustering

Cluster no 1:
['The cascade auction – a mechanism for deterring collusion in auctions', 'Optimal Coalition Structures in Cooperative Graph Games', 'Posted Prices Exchange for Display Advertising Contracts', 'Modelling and Control of Mixed Observability Multiagent Systems', 'Multiagent Learning with a Noisy Global Reward Signal', 'Ties Matter: Complexity of Manipulation when Tie-breaking with a Random Vote', 'How to Cut a Cake Before the Party Ends', 'Solving security games on graphs via marginal probabilities', 'Sample Complexity and Performance Bounds for Non-parametric Approximate Linear Programming', 'PAC Optimal Exploration in Continuous Space Markov Decision Processes', 'Computational Aspects of Nearly Single-Peaked Electorates', 'On the Value of using Group Discounts under Price Competition', 'Bundling Attacks in Judgment Aggregation', 'Bounding the Cost of Stability in Games over Interaction Networks', 'Strategic Behavior when Allocating Indivisible Goods Sequentially', 'Multi-Armed Bandit with Budget Constraint and Variable Costs', 'Equilibria of Online Scheduling Algorithms', 'Bribery in Voting With Soft Constraints', 'Interdependent Multi-Issue Negotiation for Energy Exchange in Remote Communities', 'Dynamic Social Choice: Foundations and Algorithms', 'Information Sharing Under Costly Communication in Joint Exploration', 'Cost-Optimal Planning by Self-Interested Agents', 'How Bad is Selfish Voting?', 'Efficient evolutionary dynamics with extensive-form games', 'A Framework for Aggregating Influenced CP-nets and its Resistance to Bribery', 'Search More, Disclose Less', 'A Robust Bayesian Truth Serum for Non-binary Signals', 'On the Social Welfare of Mechanisms for Repeated Batch Matching', 'Automating Collusion Detection in Sequential Games', 'Instructor Rating Markets', 'Fast Equilibrium Computation for Infinitely Repeated Games', 'Composition Games for Distributed Systems: the EU Grant games', 'Structured Kernel-Based Reinforcement Learning', 'Multi-Cycle Query Caching in Agent Programming', 'Pruning for Monte Carlo Distributed Reinforcement Learning in Decentralized POMDPs', 'Analyzing the effectiveness of adversary modeling in security games', 'Algorithms for strong Nash equilibrium with more than two agents', 'Decoupling the Multiagent Disjunctive Temporal Problem', 'Online Lazy Updates for Portfolio Selection with Transaction Costs', 'Teamwork with Limited Knowledge of Teammates']

Cluster no 2:
['Learning Integrated Symbolic and Continuous Action Models for Continuous Domains', 'Basis Adaptation for Sparse Nonlinear Reinforcement Learning', 'Reciprocal Hash Tables for Nearest Neighbor Search', 'Symmetry-Aware Marginal Density Estimation', 'A Generalized Student-t Based

Approach to Mixed-Type Anomaly Detection', 'Scalable Lifelong Learning with Active Task Selection', 'Continuous Conditional Random Fields for Efficient Regression in Large Fully Connected Graphs', 'Rank Aggregation via Low-rank and Structured-sparse Decomposition', 'Formalizing Hierarchical Clustering as Integer Linear Programming', 'Clustering with Complex Constraints - Algorithms and Applications', 'A Maximum K-Min Approach for Classification', 'Sparse Multi-task Learning for Detecting Influential Nodes in an Implicit Diffusion Network', 'Lazy Gaussian Process Committee for Real-Time Online Regression', 'Unsupervised Cluster Matching via Probabilistic Latent Variable Models', 'Projected Group Sparse Coding for Image Classification', 'Deep Manifold Learning', 'Robust Discrete Matrix Completion', 'Spectral Rotation vs K-means in Spectral Clustering', 'Guiding Scientific Discovery with Explanations', 'Smart Multi-task Bregman Clustering and Multi-task Kernel Clustering', 'A Cyclic Weighted Median Method for L1 Low-Rank Matrix Factorization with Missing Entries', 'From Interest to Function: Location Estimation in Social Media', 'Supervised Nonnegative Tensor Factorization with Maximum-Margin Constraints', 'Supervised Coupled Dictionary Learning with Group Structures for Multi-modal Retrieval', 'On Power law kernels, corresponding Reproducing Kernel Hilbert Space and applications', 'Vector-valued Multi-view Semi-supervised Learning for Multi-label Image Classification', 'Multi-Label Learning with PRO Loss', 'Towards Cohesive Anomalies Mining', 'SMILe: Shuffled Multiple-Instance Learning', 'Large-Scale Hierarchical Classification via Stochastic Perceptron', 'Teaching Classification Boundaries to Humans', 'Optimizing Objective Function Parameters for Strength in Computer Game-Playing', 'HC-Search: Learning Heuristics and Cost Functions for Structured Prediction', 'A Concave Conjugate Approach for Nonconvex Penalized Regression with the MCP Penalty', 'Convex Subspace Representation Learning from Multi-view Data', 'Time-dependent Trajectory Regression on Road Networks via Multi-Task Learning', 'A Tensor-Variate Gaussian Process for Classification of Multidimensional Structured Data', 'Discovering hierarchical structure for sources and entities', 'Uncorrelated Lasso']

Cluster no 3:
['Greedy or Not? Local search move selection for MAXSAT', 'Truncated LPA* : Faster Replanning by Exploiting Suboptimality', 'A General Formal Framework for Pathfinding Problems with Multiple Agents using Answer Set Programming', 'External Memory Best-First Search for Multiple Sequence Alignment', 'Resolution and Parallelizability: Barriers to the Efficient Parallelization of SAT Solvers', 'Improving WalkSAT for Random k-Satisfiability Problem with $k>3$', 'Abstract Preference Frameworks — a Unifying Perspective on Separability and Strong Equivalence', 'A Robust Bidirectional Search Using Heuristic Improvement', 'Improved Optimal Search Heuristics with Manifold Learning', 'Improving the Performance of Consistency Algorithms by Localizing and Bolstering Propagation in a Tree Decomposition', 'Domain-specific Heuristics in Answer Set Programming', 'Simple Temporal Problems with Taboo Regions', 'When is Brute-Force Avoidable for CSP?', 'Red-Black Relaxed Plan Heuristics', 'Temporal Milestones in HTNs', 'Liberal Safety for Answer Set Programs with External Sources', 'Backdoors to Normality for Disjunctive Logic Programs', 'A Morphogenetically Assisted Design Variation Tool', 'Extending STR to a Higher-Order Consistency', 'Mixed Heuristic Local Search for Protein Structure Prediction', 'Radial Restraint: A Semantically Clean Approach to Bounded Rationality for Logic Programs', 'Partial MUS Enumeration']

Cluster no 4:

['The Automated Acquisition of Suggestions from Tweets', 'Vesselness Features and the Inverse Compositional AAM for Robust Face Recognition using Thermal IR', 'Effective Bilingual Constraints for Semi-supervised Learning of Named Entity Recognizers', 'Gradient Networks for Shape-Based Object Instance Detection', 'Video Saliency Detection via Dynamic Consistent Spatio-Temporal Attention Modelling', 'Enforcing Meter in Finite-Length Markov Sequences', 'Automatic Identification of Conceptual Metaphors With Limited Knowledge', 'Probabilistic Sense Sentiment Similarity through Hidden Emotions', 'Joint Object and Pose Recognition using Homeomorphic Manifold Analysis', 'A Pattern Matching Based Graphical Model for Question Subjectivity Prediction', 'Grounding Natural Language References to Unvisited and Hypothetical Location', 'Story Generation with Crowdsourced Plot Graphs', 'A Hierarchical Aspect-Sentiment Model for Online Reviews', 'Generating Natural-Language Video Descriptions Using Text-Mined Knowledge', 'Incremental Learning Framework for Indoor Scene Recognition', 'A Topic-Based Coherence Model for Statistical Machine Translation', 'Joint inference of extraction and labelling via graph propagation for dictionary construction', 'Salient Object Detection via Low-Rank and Structured Sparse Matrix Decomposition', 'Integrating Programing by Example and Natural Language Programing']

Cluster no 5:
['Ranking Scientific Articles by Exploiting Citations, Authors, Journals and Time Information', 'Walking on Minimax Paths for k-NN Search', 'Social Rankings in Human-Computer Committees', 'Unified Constraint Propagation on Multi-View Data', 'Automated Workflow Synthesis', 'Sensitivity of diffusion dynamics to network uncertainty', 'An Agent Design for Repeated Negotiation and Information Revelation with People']

Cluster no 6:
['m-Transportability: Transportability of Causal Effects from Multiple Environments', 'Progression of Decomposed Situation Calculus Theories', 'Answering Counting Aggregate Queries over Ontologies of DL-Lite Family', 'Causal Transportability with Limited Experiments', 'Boosting Lifted Likelihood Maximization for MAP Inference by Virtual Evidence', 'Complexity of Inferences in Polytree-shaped Semi-Qualitative Probabilistic Networks', 'GiSS: Combining Gibbs Sampling and SampleSearch for Inference in Mixed Probabilistic and Deterministic Graphical Models', 'RockIt: Exploiting Parallelism and Symmetry for MAP Inference in Statistical Relational Models', 'Multi-agent Knowledge and Belief Change in the Situation Calculus', 'Dynamic Minimization of Sentential Decision Diagrams', 'Filtering with Logic Programs and its Application to General Game Playing']

Cluster no 7:
["Reasoning about Conditional Independence under Uncertainty: Axioms, Algorithms and Levesque's Situations to the Rescue"]

Cluster no 8:
['Data-Parallel Computing Meets STRIPS', 'Qualitative Planning under Partial Observability in Multi-Agent Domains', 'Parameterized Complexity Results for Case-Based Planning', 'A Fast Pairwise Heuristic for Planning under Uncertainty', 'Timelines with Uncontrollability', 'A Kernel Density Estimate-based approach to Component Goodness Modeling', 'Model-Lite Case-Based Planning', 'A First-Order Formalization of Commitments and Goals', 'Assumption-Based Planning: Generating plans and explanations under incomplete knowledge', 'Hypothesis Exploration for Malware Detection using Planning']

Cluster no 9:
['An Extended GHKM Algorithm for Inducing Lambda-SCFG']

CONCLUSION

Due to certain non deterministic approaches in the algorithm, different qualities of clusters are generated due to certain unavoidable differences in methodologies, such as the position of placement of the newly generated node. But the real reason boils down to the fact that which pair of clusters are to be merged when two or more pairs have the highest proximity is undefined. Hence, if a definition is established regarding the resolution of this conflict, a deterministic result might be expected. Also, the NMI against threshold for the Jaccard coefficient does not seem to follow any specific pattern and hence needs to be established using validation, as done in this experiment.

REFERENCES

1. Niwattanakul, Suphakit, et al. "Using of Jaccard coefficient for keywords similarity." *Proceedings of the international multiconference of engineers and computer scientists*. Vol. 1. No. 6. 2013.
2. Johnson, Stephen C. "Hierarchical clustering schemes." *Psychometrika* 32.3 (1967): 241-254.
3. Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." *Proceedings of the national academy of sciences* 99.12 (2002): 7821-7826.