Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems | School of Computing and Information Systems

5-2019

Clustering and its extensions in the social media domain

Lei MENG

Ah-hwee TAN
Singapore Management University, ahtan@smu.edu.sg

Donald C. WUNSCH

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons, and the Software Engineering Commons

Citation

MENG, Lei; TAN, Ah-hwee; and WUNSCH, Donald C.. Clustering and its extensions in the social media domain. (2019). *Adaptive Resonance Theory in Social Media Data Clustering*. 15-44.
Available at: https://ink.library.smu.edu.sg/sis_research/6061

This Book Chapter is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Chapter 2
Clustering and Its Extensions in the Social Media Domain

Abstract This chapter summarizes existing clustering and related approaches for the identified challenges as described in Sect. 1.2 and presents the key branches of social media mining applications where clustering holds a potential. Specifically, several important types of clustering algorithms are first illustrated, including clustering, semi-supervised clustering, heterogeneous data co-clustering, and online clustering. Subsequently, Sect. 2.5 presents a review on existing techniques that help decide the value of the predefined number of clusters (required by most clustering algorithms) automatically and highlights the clustering algorithms that do not require such a parameter. It better illustrates the challenge of input parameter sensitivity of clustering algorithms when applied to large and complex social media data. Furthermore, in Sect. 2.6, a survey on several main applications of clustering algorithms to social media mining tasks is offered, including web image organization, multi-modal information fusion, user community detection, user sentiment analysis, social event detection, community question answering, social media data indexing and retrieval, and recommender systems in social networks.

2.1 Clustering

Clustering, aimed at identifying natural groupings of a dataset, is a commonly used technique for statistical data analysis in many fields, such as machine learning, pattern recognition, image and text analysis, information retrieval, and social network analysis. This section presents a literature review on the important clustering techniques for multimedia data analysis in terms of different theoretical basis. To gain a systematical understanding on the clustering taxonomy, please look at past efforts [168, 169].

2.1.1 K-Means Clustering

K-means clustering [109] is a centroid-based partitional algorithm, which partitions the data objects, represented by feature vectors, into k clusters. It iteratively seeks k cluster centers in order to minimize the intra-cluster squared error, defined as

© Springer Nature Switzerland AG 2019

L. Meng et al., *Adaptive Resonance Theory in Social Media Data Clustering*, Advanced Information and Knowledge Processing,
https://doi.org/10.1007/978-3-030-02985-2_2

$$arg\min_{\mathscr{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in s_i} ||\mathbf{x} - \mu_i||, \tag{2.1}$$

where $\mathscr{S} = \{s_1, \ldots, s_k\}$ is a partition of data in k groups, $\mathbf{x}$ is the feature vector of a data object in the cluster s_i, μ_i is the weight vector (in this case the mean vector) of all feature vectors of cluster s_i, and $||.||$ is the vector norm, typically the Euclidean norm, measuring the distance between $\mathbf{x}$ and μ_i.

K-means clustering is widely used due to its easy implementation, linear time complexity of $O(n)$, and well-founded objective function, and many variations have been proposed such as Fuzzy C-means Clustering [123] and Kernel K-means Clustering [51]. However, it suffers from two fundamental drawbacks: (1) the number of clustering k is difficult to determine, and (2) the clustering result is sensitive to the initialization of cluster centers. Accordingly, numerous research efforts have been conducted to tackle these problems, such as [7, 9, 58].

Although such problems are still unsolved, the standard K-means clustering, in practice, frequently finds reasonable solutions quickly and is widely used in various applications, such as image segmentation [40], image organization [24], and graph theoretic clustering [136, 146].

2.1.2 Hierarchical Clustering

Hierarchical clustering algorithms attempt to generate a hierarchy of clusters for data objects. Typically, hierarchical clustering techniques fall into two types:

- **Agglomerative clustering**: Each data object is a leaf cluster of the hierarchy, and pairs of clusters are merged iteratively according to certain similarity measures.
- **Divisive clustering**: All the data objects start in one cluster, and cluster splitting is performed recursively according to some dissimilarity measures.

The agglomerative clustering algorithms typically merge leaf data objects and clusters using a combination of distance metrics for single data objects and clusters. The distance between data objects is usually defined as vector norms, such as

- Euclidean distance (ℓ_2 norm): $d(a, b) = \sqrt{\sum_i (a_i - b_i)^2}$
- Manhattan distance (ℓ_1 norm): $d(a, b) = \sum_i |a_i - b_i|$

While the similarity or dissimilarity/distance between clusters is usually measured by the linkage criteria, such as

- Single-linkage [66]: $\min\{\, d(a, b) : a \in A,\ b \in B \,\}$
- Average-linkage [144]: $\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b)$
- Complete-linkage [3]: $\max\{\, d(a, b) : a \in A,\ b \in B \,\}$

In contrast to the iterative merging of pairs of leaf and intermediate nodes, the divisive clustering usually produces each layer of the hierarchy using some clustering algorithms, such as K-means, decision tree and ART [13].

Although hierarchical clustering has been widely used in image and text domains [76, 140], three major problems remain: (1) High time complexity, usually of $O(n^3)$, limits its scalability for big datasets; (2) The generated hierarchy can be very complex for a dataset containing diverse contents; and (3) Deciding the stop criteria is difficult.

In recent years, some hierarchical clustering algorithms have been developed for web image organization [24, 53], which successively use different types of features, such as textual and visual features, to build a multi-layer hierarchy. However, this approach cannot provide a semantic hierarchy of clusters. Also, it suffers from the problem of error propagation, because the clustering result of data objects in one layer is based on that of the previous layers.

2.1.3 Graph Theoretic Clustering

Graph theoretic clustering models the relations between data objects by a graph structure where each data object is a vertex, and an edge between a pair of vertices indicates their relation. This approach is intended to group the vertices into clusters according to some optimization criteria. Graph theoretic clustering is widely studied in the literature because of its well-defined objective functions which can be easily utilized to formulate a wide range of clustering problems.

Spectral clustering, one of the most well-known graph theoretic clustering methods, refers to a type of clustering technique, such as normalized cut [146] and minimum cut [166]. For example, given an affinity matrix $\mathbf{A} = \{A_{ij}\}$ where A_{ij} is the distance between the i-th and j-th objects, the normalized cut algorithm first computes its Laplacian matrix

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \tag{2.2}$$

where $\mathbf{D}$ is a diagonal matrix where $D_{ii} = \sum_j A_{ij}$. Subsequently, the eigenvalues λ and eigenvectors $\mathbf{y}$ are obtained by solving

$$(\mathbf{D} - \mathbf{W})\mathbf{y} = \lambda\mathbf{D}\mathbf{y}. \tag{2.3}$$

By representing data objects using the first k eigenvectors, the data clusters are obtained by performing the K-means algorithm on the new data representation matrix.

To decrease the computation cost and avoid the effect caused by different similarity measures, bipartite spectral graph partitioning [135] is proposed. It directly models the relations between data and features using a bipartite graph and finds the solution by solving a singular value decomposition problem [65]. A similar idea has been applied on the image domain [133].

2.1.4 Latent Semantic Analysis

Latent Semantic Analysis (LSA) [49] is initially proposed to analyze the relationships between a set of documents and the words therein. Given a term-document matrix $\mathbf{X}$, LSA decomposes the matrix into three matrices via singular value decomposition (SVD) [65], defined as

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T, \tag{2.4}$$

where $\mathbf{U}$ and $\mathbf{V}^T$ are orthogonal matrices containing the singular vectors and Σ is a diagonal matrix containing the singular values. The new data representation is obtained using the k rows in $\mathbf{V}^T$ corresponding to the k largest singular values in Σ, i.e. the latent semantic space.

The key idea behind LSA is to map the high-dimensional term vectors of documents to a lower dimensional representation in a so-called latent semantic space. Analogous to spectral clustering, a traditional clustering algorithm should be employed to obtain the cluster assignment of data objects in the latent semantic space. LSA has been applied to a wide range of topics including text summarization [124, 151], face recognition [59], and image retrieval and annotation [130].

2.1.5 Non-Negative Matrix Factorization

Non-negative Matrix Factorization (NMF) [94], similar to Latent Semantic Analysis (LSA), is also a technique based on matrix factorization. In contrast, NMF iteratively decomposes the feature matrix $\mathbf{X} \in \Re^{n\times m}$ into two matrices $\mathbf{W} \in \Re^{n\times k}$ and $\mathbf{H} \in \Re^{k\times m}$ based on the objective function minimizing the reconstruction error, defined as

$$\min ||\mathbf{X} - \mathbf{W}\mathbf{H}||_F^2, \tag{2.5}$$

where $||.||_F^2$ is the squared Frobenius norm.

Contrary to spectral clustering and LSA that are equivalent to the feature reduction process, NMF derives the cluster indicator matrix $\mathbf{W}$ that directly reveals the relations between each of the n documents and a pre-defined number of clusters k (dimensions). As such, the cluster membership of each document is determined by the largest projection value among all the k dimensions. A study [170] indicates that NMF outperforms spectral methods in text document clustering in terms of both accuracy and efficiency.

Recently, a tri-factorization objective function [52] has been proposed for a general framework of data clustering, which has been extended to perform document-word co-clustering [68] and semi-supervised document clustering [45].

2.1.6 Probabilistic Clustering

Probabilistic clustering, usually referred to as mixture models, is a generative model-based approach, which uses statistical distributions to model clusters and achieves the cluster assignment of data objects by optimizing the fit between the data and the distributions. Specifically, this approach assumes that data objects are generated from a set of probabilistic distributions, so the data points in different clusters should follow different probabilistic distributions. Typically, this approach requires the user to specify the number and the functional forms of the distributions, such as the Gaussian distribution [113]. As such, the clustering process is equivalent to estimating the parameters of the probabilistic distributions.

Gaussian mixture model (GMM) is a commonly-used algorithm for probabilistic clustering, where each data object $\mathbf{x} \in \mathbf{X}$ is estimated by a weighted sum of k Gaussian distributions, defined as

$$p(\mathbf{x}|\theta) = \sum_{i=1}^{k} \phi_i \mathcal{N}(\mathbf{x}|\mu_i, \sigma_i), \tag{2.6}$$

$$s.t. \sum_{i=1}^{k} \phi_i = 1 \tag{2.7}$$

where θ is the set of parameters of distributions to be estimated, $\mathcal{N}(\mu_i, \sigma_i)$ is the i-th Gaussian distribution and ϕ_i is the corresponding weight.

The objective function of GMM is to maximize p($\mathbf{x}$) for each data object, defined as

$$arg \max_{\theta} \sum_{\mathbf{x} \in \mathbf{X}} p(\mathbf{x}) \tag{2.8}$$

The most popular method for solving the parameter estimation task of probabilistic distributions defined in Eq. (2.8) is the Expectation-Maximization (EM) algorithm [50, 121, 132], which estimates the maximum likelihood of parameters based on Bayes's theorem. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood and is evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters by maximizing the expected log-likelihood found in the E step.

2.1.7 Genetic Clustering

The use of genetic algorithms [10, 11] to identify the best clustering typically depends on the evolution of cluster structures, as evaluated by certain cluster validity indices.

As an example, the symmetry-based genetic clustering algorithm, called VGAPS-clustering [11], models a cluster structure $\mathbf{x}$ as a "chromosome" where the "genes" are a concatenation of the number of clusters and cluster weights, such that $\mathbf{x} = [j, x_1, \ldots, x_j]$. VGAPS-clustering thereafter randomly generates a number of such "chromosomes" as a population pool, and uses a fitting function $f(\mathbf{x})$ to select the best-fitting "chromosomes", defined as

$$arg \max_{\mathbf{x}} f(\mathbf{x}), \tag{2.9}$$

where the $f(\mathbf{x})$ is customized and a typical choice is that of the K-means clustering algorithm as defined in Eq. (2.1).

The selected best-fitting patterns of cluster structures are then modified to generate the next generation pool using the typical evolutionary operators of genetic algorithms, such as "mutation", "crossover" and "selection". Note that different patterns may have different numbers of centers. After the maximum number of generations, the pattern with the highest fitness is selected as the best cluster structure.

Genetic clustering can identify clusters of arbitrary shapes and achieve a global optimum. However, genetic clustering algorithms are usually quite slow due to the stochastic evolution of patterns. The experiments presented in [11] were only conducted on a few small datasets with several hundred patterns, each of which also had a small number of dimensions. A review of genetic clustering algorithms is provided in [10].

2.1.8 Density-Based Clustering

Density-based clustering identifies dense regions of data objects as clusters in the feature space. As the first density-based clustering algorithm, DBSCAN [57] forms the degree of density using two parameters, namely, the maximum distance for the search of neighbors ε and the minimum number of neighbors $minPts$; data objects and their neighbors that satisfy the above requirements are called core points and are deemed to be in the same cluster. The data objects that do not satisfy the requirements and are not neighbors of any core point are considered noise. Following the above criteria, DBSCAN examines all the data objects and identifies clusters and noise.

In addition, DBSCAN has several extensions. GDBSCAN [142] extends DBSCAN so that it can cluster point objects and spatially extended objects according to both their spatial and non-spatial attributes. OPTICS [6] provides a hierarchical view of the data structure, which is equivalent to the density-based clusterings corresponding to a broad range of parameter settings. DECODE [128] composes clusters with different densities in the dataset; more specifically, it computes the mth nearest distance of each pattern and uses reversible jump Markov Chain Monte Carlo (MCMC) to identify the clusters of patterns in terms of their different densities. Tran et al. [159] proposed a density-based clustering algorithm, KNNCLUST, in which the density is measured by a KNN-kernel table. With a pre-defined number of neighbors, all the patterns in the dataset are assigned to clusters according to the proposed KNN-kernel Bayes' class-condition. The cluster memberships of all the patterns are recalculated until their cluster assignments stop changing.

Density-based algorithms have several advantages, including their ability to form clusters with arbitrary shapes and their insensitivity to initialization. However, they require several pre-defined parameters that are difficult to decide, such as the minimum number of neighbors in DBSCAN, the value of m and the parameter for deciding the probability mixture distribution in DECODE, the number of neighbors in the KNN table and the choice of kernel in KNNCLUST. Additionally, density-based clustering algorithms typically require a quadratic time complexity of $O(n^2)$, which may be reduced to $O(n\,log\,n)$ when a spatial index structure is used to speed up the search process for neighbors [142]. A review of density-based clustering algorithms can be found in [90].

2.1.9 Affinity Propagation

Affinity Propagation [60] is an exemplar-based clustering algorithm that identifies a set of representative data objects (patterns) as "exemplars" to the other patterns in the same cluster. Exemplars are identified by recursively updating two messages of patterns, namely, the "availability" $a(i, k)$ to indicate the qualification of the k-th data object to be an exemplar of the i-th data object, and the "responsibility" $r(i, k)$ to indicate the suitability of the i-th data object to be a member of the k-th exemplars' clusters. The algorithm stops when the exemplars for all the patterns remain for a number of iterations, or upon reaching a maximum number of iterations.

Two algorithms [62, 171] have been proposed to improve the efficiency of Affinity Propagation. Fast Sparse Affinity Propagation (FSAP) [171] generated a sparse graph using the K-nearest neighbor method, rather than the original similarity matrix, to reduce the computation of message transmission in Affinity Propagation. In [62], the proposed fast algorithm for Affinity Propagation reduced the computation by pruning the edges that can be directly calculated after the convergence of Affinity Propagation.

Affinity Propagation has shown better performance than K-means in terms of the average squared error. However, it has a quadratic time complexity of $O(tn^2)$ where t is the number of iterations. Even the fastest one [62] has a quadratic time complexity of $O(n^2 + tm)$, where m is the number of edges. Additionally, Affinity Propagation usually requires the tuning of four parameters, including the preference vector "preference" which controls the number of generated clusters and impacts the speed of convergence, the damping factor "dampfact" and the maximum and minimum number of iterations "maxits" and "convits" which ensure convergence.

2.1.10 Clustering by Finding Density Peaks

Cluster$_{dp}$ [139] identifies data clusters by finding the density peaks. It does not follow traditional density-based algorithms, such as DBSCAN. Instead, with a predefined value of search radius d_c, the local density ρ_i and the nearest distance to the data objects with the higher local density δ_i for each data object is computed. The density peaks are evaluated by following two criteria: 1) Density peaks should have more neighbors than those of their neighbors; and 2) all density peaks should be far away from each other. Cluster$_{dp}$ requires human decisions to select the density peaks. It plots the "$\rho - \delta$" decision graph for the users and asks them to identify those density peaks appearing in the upper-right part of the graph that are a sufficient distance from the other points. These density peaks will serve as cluster centers, and the remaining patterns are assigned to the nearest cluster centers.

This approach is fast (in time complexity of $O(n)$) and roughly robust to the single parameter d_c. However, it was found to be ineffective in the identification of representative peaks for social media data, mainly due to the high dimensionality of data and the noise in many ways [116].

2.1.11 Adaptive Resonance Theory

Adaptive Resonance Theory (ART) [30, 67] is a learning theory on how a human brain memorizes events and objects, and it leads to a series of real-time unsupervised learning models capable of fast and stable category recognition, such as ART 1 [180], ART 2 [28], ART 2-A [33], ART 3 [29], and Fuzzy ART [34], as well as supervised learning models, such as ARTMAP [32] and Fuzzy ARTMAP [31].

The ART-based clustering algorithms have different learning operations but follow similar procedures, which incrementally perform real-time searching and matching between input patterns (data objects) and existing clusters (memory prototypes) in the category space one at a time. Specifically, given an input pattern **x**, ART performs the following actions

1. Searching for the best-matching (winner) cluster c_j in the category field using a choice function $T(\mathbf{x}; c_j)$.
2. If c_j exists, a match function $M(\mathbf{x}; c_j)$ is used to determine if the degree of matching reaches a threshold, called the vigilance parameter ρ.
3. Satisfying the vigilance criteria leads to a "resonance", i.e. the input pattern is assigned to the winner c_j. Otherwise, winners in the remaining clusters are selected one by one for Step 2 until one of them passes the vigilance criteria or all of them are presented.
4. If resonance occurs, the winner cluster c_j updates its weight vector. Otherwise, a new cluster is generated to encode the input pattern.

ART has advantages of fast and stable learning as well as an incremental clustering manner, and it has been successfully applied to many applications, such as pattern recognition and document organization [112]. However, since ART achieves stable learning by depressing the values of weight vectors if the intra-cluster data objects have varied values in the corresponding features, it may suffer from the problem of "category proliferation". That is, a cluster's weight values may approach 0's after learning from ill-represented data objects.

The above problem is addressed by Fuzzy ART with the incorporation of fuzzy operators and complement coding. The use of fuzzy operators replaces the intersection operator ($\cap$) used in ART 1 with the min operator ($\wedge$) used in fuzzy set theory; while the complement coding concatenates the input feature vector $\mathbf{x}$ with its counterpart $\bar{\mathbf{x}} = \mathbf{1} - \mathbf{x}$ (Note that ART requires input values to be in [0, 1]). These changes enable Fuzzy ART to normalize the input patterns and limit the size of the clusters. More importantly, Section 3.2 illustrates how they change the clustering mechanism of ART 1 to prevent category proliferation.

Fuzzy ART has been used in different ART-based variants to resolve many image and text mining problems, such as web document management [156], tag-based web image organization [114], image-text association analysis [82], multimedia data co-clustering [117] and social community detection in heterogeneous social networks [115]. Related case studies will be presented in Part II.

2.2 Semi-Supervised Clustering

Clustering organizes data objects into groups according to purely similarity (or distance) measures in the feature space, whereas semi-supervised clustering exploits the available prior knowledge, also called side information, to guide the clustering process. Typically, the prior knowledge is given by the information about the related and/or irrelevant data objects. Group label constraint and pairwise constraint are two commonly used methods for providing such information.

2.2.1 Group Label Constraint

Group label constraint requires users to indicate subsets of documents in the dataset that belong to the same class. Semi-supervised learning is usually achieved by learning a metric for adjusting the similarity measure [145, 167] or incorporating such constraints to adjust the objective function of the original clustering algorithms [84, 108]. This type of constraint usually has conditions on the size of the subsets for performance improvement.

2.2.2 Pairwise Label Constraint

Pairwise Label constraint is the most widely used method in practice because it is easily accessible to users and does not require them to have much prior knowledge. Using this method, users need to provide a set of must-link and cannot-link constraints to indicate if pairs of documents should be associated with the same cluster or not. Chen et al. developed two methods for incorporating the pairwise constraints into the Non-negative Matrix Tri-Factorization (NMF) algorithm [52]. The first method [45] adds the constraints into the objective function as rewards and penalties to balance the clustering. The other method [44] computes new relational matrices for documents through a distance metric learning algorithm such that, in the derived feature space, documents with must-link are moved closer while those with cannot-link are moved farther apart. Besides the NMF, spectral constrained clustering algorithms for incorporating pairwise constraints have also been widely studied [61, 147]. Other notable works include Semi-supervised Kernel K-means (SS-KK) [91] and Semi-supervised Spectral Normalized Cuts (SS-SNC) [80].

2.3 Heterogeneous Data Co-Clustering

Heterogeneous data co-clustering, also called high-order multiview/multimodal clustering, addresses the problem of clustering composite objects, which are described by the data from heterogeneous resources. Typically, the data objects, such as images, text documents, and social users, and their associated descriptive information are modeled as a star structure [47]. By simultaneously integrating those different types of data as the multi-modal features of the composite objects, the heterogeneous data co-clustering task is to find the best partitioning of the composite objects, considering their similarities in terms of each feature modality.

This section illustrates existing heterogeneous data co-clustering algorithms in terms of different model formulations, which can be organized into six categories, discussed as follows.

2.3.1 Graph Theoretic Models

A large body of recent literature on heterogeneous data co-clustering is based on graph theoretic models. Gao et al. [63] proposed a web image co-clustering algorithm, named Consistent Bipartite Graph Co-partitioning (CBGC). This algorithm interprets the image-text co-clustering task as a tripartite graph and transforms the partitioning of the tripartite graph into the simultaneous partitioning of the visual and textual graphs. In this way, CBGC models the solution as a multi-objective optimization problem which is solved by semi-definite programming (SDP). This work has been generalized to process multimodal heterogeneous data in [64]. However, CBGC requires empirical settings of three parameters, and it should employ traditional clustering algorithms on the embedding vectors produced to obtain the final clusters.

A similar work [136] to CBGC, called) Consistent Isoperimetric Highorder Co-clustering (CIHC), also considers the problem of integrating visual and textual features as the partitioning of a tripartite graph. Contrary to CBGC, CIHC solves the problem by extending the Isoperimetric Co-clustering Algorithm (ICA) [135], which can be solved by a sparse system of linear equations. CIHC has been demonstrated to be more effective and has a much lower time cost than CBGC. However, it also requires an additional clustering algorithm to partition the obtained embedding vectors, and it is only applicable for distinguishing data of two classes.

Long et al. [110] proposed Spectral Relational Clustering (SRC) for clustering multi-type relational data. They first proposed a collective clustering based on minimizing the reconstruction error of both the object affinity matrices and the feature matrices, and then they derived an iterative spectral clustering algorithm accordingly for the factorization of these relational matrices. However, SRC requires solving the eigenvalue decomposition problem which is inefficient for large-scale datasets. Moreover, a separate clustering algorithm, in this case K-means, is used to obtain the final clustering.

Zhou et al. [181] proposed a multi-view spectral algorithm for clustering data with multiple views. This method generalizes the normalized cut from a single view to multiple views by forming a mixture of Markov random walks on each graph, and it aims to divide the data objects into two clusters, which should be a good partitioning for each of the graphs. Therefore, this method is not suitable for clustering datasets with many underlying clusters.

Cai et al. [25] proposed Multimodal Spectral Clustering (MMSC) to simultaneously integrate five types of visual features for image clustering. In order to obtain the final cluster indicator matrix, MMSC uses a unified objective function to simultaneously optimize the clustering results of each feature modality and their combination. This objective function is finally solved by eigenvalue decomposition and a spectral rotation algorithm.

Multimodal Constraint Propagation (MMCP) [61] has been proposed for the semi-supervised clustering of multi-modal image sets. MMCP first defines the random walk on multiple graphs, each of which corresponds to one type of modality. Subsequently, by decomposing the problem of label propagation on multiple graphs into a set of independent multi-graph-based two-class label propagation sub-problems, MMCP deduces the refined similarity matrix of data objects through a series of quadratic optimization procedures. A spectral clustering algorithm is applied to obtain the final clustering results.

In view of the above issues, the graph theoretic models typically utilize a unified objective function to realize the fusion of multi-modal features, and they require a series of matrix operations to deduce a vector or matrix that reveals the features of data objects. It is notable that the graph theoretic models deal with the similarity matrix of the data objects instead of the feature matrix. So, in practice, evaluating the similarities between data objects should be considered first. A drawback of this approach is the computational complexity due to the mathematical computation. Also, the clustering performance depends on the traditional clustering algorithms that are used to obtain the final results.

2.3.2 Non-Negative Matrix Factorization Models

The non-negative matrix tri-factorization (NMF) approach, as illustrated in Sect. 2.1.5, iteratively factorizes the data matrix into three sub-matrices. One of these, called the cluster indicator matrix, reveals the projection values of data objects to the dimensions (clusters).

Chen et al. [47] proposed a symmetric nonnegative matrix tri-factorization algorithm, called Semi-Supervised NMF (SS-NMF), which attempts to find a partitioning of the data objects to minimize the global reconstruction error of the relational matrices for each type of data. Like the NMF, the cluster membership of each data object is determined by the largest projection value among all clusters. Moreover, by incorporating the user-provided pairwise constraints, SS-NMF derives new relational matrices through a distance learning algorithm to enhance the clustering performance.

Linked Matrix Factorization (LMF) [158] has an objective function similar to that of SS-NMF. However, LMF minimizes the overall reconstruction error and maximizes the sparsity of the factorized sub-matrices at the same time. Also, a semi-supervised version using pairwise constraints is proposed for metric learning.

The NMF approach has the advantage of a linear time complexity of $O(tn)$, where t is the number of iterations and n is the number of data objects. However, it requires users to set the number of clusters for the data objects and each type of features to construct the sub-matrices, and its performance may vary with different initializations of the sub-matrices.

2.3.3 Markov Random Field Model

Bekkerman et al. [18] proposed Combinatorial Markov Random Fields (Comrafs) for co-clustering multimodal information based on the information bottleneck theory, and applied it to various applications, such as semi-supervised clustering [17], multimodal image clustering [16] and cluster analysis [19].

Comrafs constructs a set of Markov random fields for each type of data, wherein each data modality is modeled as a combinatorial random variable which takes values from all the possible partitions, and the edges between pairs of variables are represented using mutual information. The approach of Comrafs is to maximize the information-theoretic objective function, which is resolved by the hierarchical clustering algorithm with either agglomerative or divisive strategies.

One potential problem with this approach is the heavy computational cost, having a time complexity of $O(n^3 \log n)$. As Comrafs needs to traverse all subsets of the data samples for each data modality, the computational cost increases significantly with respect to the size of datasets.

2.3.4 Multi-view Clustering Models

The multi-view clustering models consider the clustering of data objects with two types of features. Typically, two clustering algorithms are employed for each set of features. Subsequently, the learned parameters of two clustering models are refined by learning from each other iteratively. However, this approach is restricted to two types of data.

In [21], three types of traditional clustering algorithms, namely Expectation-Maximization (EM), K-means and agglomerative clustering algorithms, are extended to fit the multi-view clustering framework. Additionally, the extended EM and K-means algorithms have been applied for discovering communities in linked data [55].

Recent studies also developed multi-view clustering models based on Canonical Correlation Analysis [39] and spectral clustering [92].

2.3.5 Aggregation-Based Models

The aggregation approach follows a similar idea of first identifying the similarity between the data objects through each type of features, and subsequently integrating them to produce the final results.

Principal Modularity Maximization (PMM) [157] first obtains a fixed number of eigenvectors from the modularity matrices which are produced with each type of relational matrix. Then, those eigenvectors are concatenated into one matrix, and singular value decomposition is employed to obtain the final embedding vectors for each data object. Finally, K-means is used to obtain the final clustering.

MVSIM [22] is an iterative algorithm based on a co-similarity measure termed X-SIM, which, given a relational matrix, evaluates both the similarity between the data patterns and the features. In each iteration, MVSIM runs X-SIM on the relational matrices for each feature modality to obtain the similarity matrices and then aggregates them to form an integrated similarity matrix using an update function.

2.3.6 Fusion Adaptive Resonance Theory

As discussed in Sect. 2.1.11, Adaptive Resonance Theory (ART) is an incremental clustering algorithm. It processes input patterns one at a time and employs a two-way similarity measure for the real-time searching and matching of suitable clusters to the input patterns.

Fusion ART [155] extends ART from a single input field to multiple ones and learns multi-channel mappings simultaneously across multi-modal pattern channels in an online and incremental manner. As a natural extension of ART, Fusion ART is composed of multiple input feature channels, each of which corresponds to one type of features. Thus, each type of features of the data objects is processed independently, and the output similarities of each feature channel are integrated through a choice function.

Contrary to existing heterogeneous data co-clustering algorithms, Fusion ART allows the flexibility of using different learning methods for different types of features, and it considers both the overall similarity across feature channels and the individual similarity of each modality. More importantly, Fusion ART has a very low computational complexity of $O(n)$, so it is suitable for clustering large-scale datasets. Successful applications in the multimedia domain [82, 122] have demonstrated the viability of Fusion ART for the multimedia data analysis.

2.4 Online Clustering

The large-scale and high-velocity nature of social media data, especially the streams of user-generated content, raises the need of online learning capability for clustering algorithms. It enables clustering algorithms to perform real-time processing, learn from input data objects one at a time and evolve the structure of data clusters without re-visiting past data.

2.4.1 Incremental Learning Strategies

Incremental clustering [1, 4, 12, 38, 69] belongs to a more general class, called stream clustering [148], which has attracted attention for decades. It is a special case of online learning, and it aims to enable one- or several-pass processing of the dataset one by one or in small batches instead of the whole for the purpose of saving time and memory cost.

2.4.2 Online Learning Strategies

Online clustering [46, 105, 120, 148, 175] is another branch of stream clustering. Beyond incremental clustering that clusters static data, it incorporates the online learning property that allows not only incrementally processing but also continuous learning from streaming data. However, existing algorithms in the literature usually are k-means or hierarchical clustering variants requiring the specification of either the number of clusters or more than two parameters. As illustrated in Section 1.2.5, this affects the robustness of these algorithms for large-scale and noisy social media data and makes human intervention intractable.

2.5 Automated Data Cluster Recognition

Existing clustering algorithms typically require setting the number of clusters in datasets. However, contrary to traditional image and text document datasets, social media data is usually large-scale and may cover diverse content across different topics, making it difficult to manually evaluate the number of underlying topics in the datasets. Therefore, automatically identifying the number of clusters in datasets becomes a key challenge for clustering social media data.

This section introduces existing approaches on the automatic recognition of clusters in a dataset.

2.5.1 Cluster Tendency Analysis

Cluster tendency analysis aims to identify the number of clusters in a dataset before clustering. Most recent studies [20, 149, 161] have focused on investigating the dissimilarity matrix of patterns.

Visual Assessment of Tendency (VAT) [20] reorders the dissimilarity matrix of patterns so that patterns in nearby rows will have low dissimilarity values. When displaying the reordered matrix as an intensity image, referred to as a "reordered dissimilarity image" (RDI), the number of clusters may be determined by counting the dark blocks along the diagonal pixels in the image. However, in complex datasets, the boundaries between dark blocks may be indistinct, making it difficult to correctly identify the number of clusters.

Therefore, Cluster Count Extraction (CCE) [149] and Dark Block Extraction (DBE) [161] are further proposed to objectively identify the number of clusters without relying on manual counting. CCE attempts to remove noise in the RDI obtained by VAT through two rounds of Fast Fourier Transform (FFT) with a filter that transforms the RDI to and from the frequency domain. The number of clusters equals the number of spikes in the histogram constructed by the off-diagonal pixel values of the filtered image. In contrast, after obtaining the RDI, DBE employs several matrix transformation steps to project all the pixel values of the RDI to the main diagonal axis to obtain a projection signal. The number of clusters equals the number of major peaks in the signal.

In practice, a traditional clustering algorithm, such as K-means, can be employed to obtain the clusters using the identified number of clusters. However, such methods have several limitations when applied to web multimedia data. First, because these datasets typically involve noise, the dissimilarity matrix may not represent the structure of the data in the input space well, which may result in an RDI with low quality. Second, such methods employ heavy computation, so their performance is only measured on small datasets containing several thousand patterns.

2.5.2 *Posterior Cluster Validation Approach*

Cluster validation aims to quantitatively evaluate the quality of different cluster structures, usually based on intra-cluster compactness and between-cluster separation, to find the best clustering.

Liang et al. [103] proposed a modified K-means algorithm with a validation method based on the intra-cluster and between-cluster entropies. This algorithm requires K-means to run multiple times, starting with a pre-defined maximum number of clusters. During each iteration, the "worst cluster" is removed using information entropy, and the quality of the clusters is evaluated according to the proposed validation method. Upon reaching the pre-defined minimum number of clusters, the clustering with the best quality is identified.

In [153], Sugar et al. proposed a "jump method", which generates a transformed distortion curve based on the clustering results of K-means with different numbers of clusters. The highest peak, or "jump", in the curve represents the best number of clusters.

Kothari et al. [89] proposed a scale-based algorithm in which a "neighborhood" serves as the scale parameter. By varying the value of the neighborhood, the proposed algorithm may identify clusterings with different numbers of clusters. The best number of clusters is identified based on the persistence across a range of neighbors.

A meta-learning-based algorithm was proposed in [95]. Given a dataset, multiple subsets are first generated by distorting the original patterns. Subsequently, for each subset, a traditional clustering method is employed to generate clusterings with different numbers of cluster; the quality of these is measured by the disconnectivity and compactness. After identifying the elbows of both the disconnectivity and the compactness plots for each subset, the true number of clusters is decided by a vote.

The above methods are typically designed for hard clustering algorithms. For fuzzy clustering algorithms, a summary of existing cluster validity indices can be found in [154, 162].

2.5.3 *Algorithms Without a Pre-defined Number of Clusters*

As discussed above, the cluster tendency analysis requires heavy computation and is not robust to noise. Similarly, the cluster validation approach attempts to select the best number of clusters by evaluating the quality of clusterings with different numbers of clusters. As such, they are not feasible for the large-scale social media datasets.

Fortunately, there are clustering algorithms that do not require a pre-defined number of clusters, including the hierarchical-clustering-based algorithms, genetic clustering algorithms, density-based clustering algorithms, Affinity Propagation and ART-based clustering algorithms. The hierarchical clustering and genetic clustering algorithms, especially, are theoretically similar to the cluster validation approach, which generates different cluster structures of patterns and employs cluster validation methods to evaluate the quality of newly generated clusters to identify the best cluster structure.

As discussed in Sect. 2.1.2, hierarchical clustering algorithms either merge small clusters with individual data objects into big clusters or split the dataset into individual data objects step by step. Therefore, existing studies typically incorporate a cluster validity index to measure the cluster quality during each merging or splitting iteration. Li et al. [101] proposed an Agglomerative Fuzzy K-means algorithm that introduces a penalty term to the objective function of the standard Fuzzy K-means and requires a maximum number of clusters. The modified Fuzzy K-means runs multiple times with a gradually increased penalty parameter; during these runs, the clusters that share centers are merged according to a validation method. The algorithm stops when the number of cluster centers remains stable over a certain number of iterations. Leung et al. [98] proposed a scale-based algorithm, based on the scale space theory, in which a dataset is considered an image, and each pattern is considered a light point on the image. The generation of a hierarchy is then simulated by blurring the image such that the light points gradually merge together. Several cluster validity indices, including lifetime, compactness, isolation and outlierness, are used to select the best cluster structure in the hierarchy. In [172], an agglomerative clustering algorithm was proposed for transactional data. Based on the intra-cluster dissimilarity measure, referred to as the "coverage density", a "Merge Dissimilarity Index" is presented to find the optimal number of clusters.

Detailed illustrations of genetic clustering algorithms, density-based clustering algorithms, Affinity Propagation, and ART-based clustering algorithms can be found in Sects. 2.1.7, 2.1.8, 2.1.9, and 2.1.11 respectively.

Although the aforementioned algorithms do not require the number of clusters to be set, they employ other parameters to determine the properties of patterns in the same cluster. The advantages of ART-based algorithms over density-based clustering algorithms and Affinity Propagation include their low time complexity and the use of a single ratio value (the vigilance parameter) to form clusters.

2.6 Social Media Mining and Related Clustering Techniques

Social media data refers to data that is generated by the users on social websites, such as the tweets on Twitter, blogs published on Facebook, images shared on Flickr, questions and answers on Yahoo! answers, and user comments and descriptions for the above user-generated multimedia data.

As previously mentioned, the big social media data record user behaviors and activities on social websites and provide rich information for multimedia data understanding and social behavior analytics. However, contrary to traditional dat sets for data mining tasks, they are large scale, noisy, multimodal, unstructured and dynamic in nature, due to the diverse ways for communicating between users provided by social websites.

Therefore, those distinguishing characteristics of social media data pose new challenges for developing novel techniques to utilize the rich but noisy information for traditional multimedia data understanding and mining tasks, such as tag-based web image organization [24, 83], comment-based video organization [76], image retrieval assisted by web images and their surrounding text [42], short text understanding [77, 150] and multimodal feature integration for social media data understanding [47, 117]. Additionally, numerous new problems and requirements arise, which are important for social media research and development, such as social community discovery [8, 115, 126, 173], user sentiment analysis [107, 125], influential user detection [2, 35], social link prediction and recommendation [54, 88, 174], question answering system analysis [5, 75], and emerging social event recognition and prediction [14, 102, 141]. A brief introduction of social media mining can be found in [70].

The following sections illustrate several directions of social media mining tasks that utilize clustering techniques as a solution.

2.6.1 Web Image Organization

The vast number of web images online motivates the requirement of effective image organization, especially the search results from web engines. Due to the diverse nature of web image content, it is difficult to group images with similar semantics solely based on the visual features. Therefore, early efforts are usually based on the clustering of the textual features extracted from the surrounding text of web images [76, 83].

Additionally, there are some studies [24, 53] that make use of both the visual content and the surrounding text of web images to generate a two-layer hierarchical structure. Those methods typically apply clustering algorithms to the textual features to generate the first layer of clusters, and subsequently group the images in each cluster according to their visual features.

Besides the tag-based image organization techniques, there are also studies on improving the organization of the image search results using purely visual features. Leuken et al. [96] developed three clustering algorithms that can incorporate multiple types of visual features for partitioning images with different visual appearances. A weighting function is proposed to dynamically evaluate the distinguishing power of the algorithms.

Recently, crowdsourcing has been incorporated into the clustering techniques as a solution to improve the clustering performance of web images [43]. By asking web users to judge the cluster membership of some images, this type of clustering models utilizes such information as relevance constraint to learn a new distance metric for refining the clustering performance.

2.6.2 Multimodal Social Information Fusion

The images and text documents in social media are usually attached with rich meta-information, such as category information, user description and user comments. Multi-modal information fusion, therefore, is aimed at processing those interrelated data modalities in a unified way and identifying their underlying interactions.

Image-text fusion for image clustering is widely studied for alleviating the semantic gap [114]. Early studies attempt to integrate the visual and textual features by either concatenating them into a single vector [180] or using them in a sequential manner [24]. However, the first approach usually cannot achieve the desired results. The second method suffers from the problem of error propagation, and the sequential usage of textual and visual features does not help improve the clustering quality. Jiang et al. [82] interpret the fusion of visual and textual features as identifying pairs of related images and texts, and propose two methods, based on vague transformation [81] and Fusion ART [155], for learning the image-text associations. Existing clustering techniques in the literature for the fusion of multimodal features will be discussed in Sect. 2.3.

The fusion of multi-modal features is also an important research task for various applications, such as multi-document summarization [72, 160] and multi-modal multimedia data indexing and retrieval [36, 99, 118, 134].

2.6.3 User Community Detection in Social Networks

A user community is formed when a group of social users have similar interests or behaviors or interact with each other more frequently on the Web than those outside of the group. The user community detection task is thus to identify different underlying communities in social networks, which may further benefit relevant research tasks, such as collective social behavior analysis [177] and social link prediction and recommendation [54, 88, 174].

A social network of users is typically modeled as a graph, where each node corresponds to a user and each edge indicates the strength of the connection between two users, such as the frequency of contact or the number of co-subscription. Clustering is commonly used for the community detection task, especially the graph theoretic clustering algorithms. However, there are two challenges for applying traditional clustering algorithms to clustering social networks. The first challenge is the large-scale size of the social network data. To overcome this problem, existing studies attempt to reduce the computational cost of their algorithms by obtaining an approximate solution from the simplified network graphs [111, 143] or developing parallel clustering models [164]. In addition to the problem of big data, the other problem is the lack of ground-truth. Existing studies on assessing the quality of the discovered clusters are usually based on the internal similarities or distances between nodes. Yang et al. [173] presented a comparative study of 13 evaluation measures for discovering the densely connected users as communities.

In recent years, a large body of studies focused on discovering user communities in the heterogeneous social networks. That is, users are connected with different types of links. Some of the recent studies on this topic are based on multi-view clustering approach [55], matrix factorization approach [158] and aggregation approach [158]. Additionally, this task is closely related to heterogeneous data co-clustering, as discussed in Sect. 2.3.

2.6.4 User Sentiment Analysis

The analysis of user sentiment is aimed at understanding the users' attitudes and opinions from their comments on products, services and events.

Most of the existing studies are based on supervised learning while those based on unsupervised learning are inadequate [78, 182]. Clustering algorithms, in this task, are typically performed to identify groups of users or comments that reveal similar sentiment, such as positive, negative and neutral. Hu et al. [78] incorporated emotional signals, such as emoticons and sentiment lexicon, into a non-negative matrix tri-factorization clustering algorithm to discover groups of users with similar sentiment. Zhu et al. [182] also developed a non-negative matrix tri-factorization model for clustering user and user comments. Moreover, an online framework has been proposed to receive dynamic online streams. A review of unsupervised sentiment analysis methods can be found in [78].

2.6.5 Event Detection in Social Networks

Clustering-based social event detection aims to identify the social events that attract collective attention through the massive number of posts and comments of users on social networking websites.

There are two directions for social event detection. One type of study focuses on detecting real-time social events through online clustering algorithms. Becker et al. [14] developed an online clustering model with a set of cluster-level event features to group Twitter messages, and subsequently trained a classification model to judge whether the generated clusters are related to events.

The other type focuses on detecting social events from a set of user messages collected from a given time period, also known as retrospective event detection [41]. Chen et al. [41] utilized the tags, time stamps, and location information of the images collected from Flickr to cluster these images and simultaneously obtain the key tags of clusters as events. Papadopoulos et al. [127] developed a clustering algorithm to cluster tagged images using their visual and textual features, and subsequently used a classifier to determine whether the clusters of images represent events or landmarks. Petkos et al. [129] developed a multi-modal spectral clustering algorithm for clustering multimedia data with different attributions, such as time, location, visual features and tags.

2.6.6 Community Question Answering

The community question answering task attempts to resolve the problem of automatically providing answers to user's questions based on a question-answer database.

In this task, the user's question is typically treated as a query, and clustering is usually adopted to identify the question-answer pairs that are similar to the user query. Subsequently, answer ranking is further employed to produce relevant answers. In an early work, Kwok et al. [93] developed a question answering system, called Mulder. It first obtains a set of answers by sending the user's query to several search engines, and it then uses a clustering algorithm to group similar answers together. Finally, a voting procedure is conducted to select the best-matching answer. Blooma et al. [23] modeled the question-answer pairs as a question-answer-asker-answerer quadripartite graph and proposed an agglomerative algorithm to merge similar question-answer pairs. A review of the related question answering studies can be found in [87].

The community question answering task is also closely related to the task of query clustering [15, 97, 179], which addresses the problem of identifying and organizing similar user queries to web search engines.

2.6.7 Social Media Data Indexing and Retrieval

Multimodal image indexing and retrieval typically follow two main approaches. The first approach is to extend existing algorithms for image indexing with a single type of features for integrating multiple types of features. Examples include Latent Semantic Indexing (LSI) [27, 37], probabilistic Latent Semantic Analysis (pLSA) [37, 106], and Non-negative Matrix Factorization (NMF) [26]. Caicedo et al. [27] proposed a Latent Semantic Kernel (LSK), based on LSI, which adopts kernel methods to compute the similarity between the query and the indexed images. Multimodal LSI (MMLSI) [37] utilizes tensors for multimodal image representation and employs Higher Order Singular Value Decomposition (HOSVD) [48] for obtaining the feature representation of images. Chandrika et al. [37] extended pLSA by jointly considering visual and textual features in a probabilistic model and employed the EM algorithm to obtain the derived representation of the images. The Multilayer Multimodal probabilistic Latent Semantic Analysis (MM-pLSA) [106] handles the visual and textual information of images via a multi-layer model, which consists of two leaf pLSA models for learning the visual and textual representation of images respectively, and a node pLSA for obtaining a unified representation. Caicedo et al. [26] proposed two methods based on Non-negative Matrix Factorization (NMF), of which the first method concatenates the matrices for visual and textual features in order to enable searching through both visual and textual features, while the second method aims to successively optimize the transformation matrices of textual and visual features, which enables searching by using either visual features or keywords.

The second approach is to construct a new representation by exploring the association among multimodal features. Li et al. [100] proposed four methods to infer the similarity matrices for the visual and textual features. The learned similarities are utilized for tackling image retrieval based on visual or textual features. Escalante et al. [56] proposed two methods for image indexing based on the occurrences and co-occurrences information of terms in the surrounding text and the object labels associated to images. The hybrid framework [152], named iSMIER, performs image retrieval by predicting the captions and annotations for the query image and indexing it by its visual fuzzy membership of clusters.

2.6.8 Multifaceted Recommendation in Social Networks

Recommendation is one of the most important techniques in the era of the social Web in both academic and industrial domains. Utilizing the rich multimedia data online, the recommendation techniques will be able to analyze the information (E-commerce products, mobile apps, or new friends online) to be diffused and targeted towards suitable populations. Some popular directions include location-based recommendation, online-offline recommendation, and explicit/implicit feedback-based recommendation. To be concrete, given an e-commerce product coupon, the recommendation algorithm will analyze the profiles of the people who are likely to purchase this product. These profiles may be the users' living and shopping locations, users' online browsing and purchase records, users' interactions and feedbacks to similar products, or the similarity of the users' interests to other users who may be willing to buy such products.

The above scenarios lead to the need for multifaceted recommendation [138], where information on the profiles of users from different sources can be gathered and analyzed for effective recommendation. This task is commonly addressed using the collaborative filtering approach,[1] which is a general term describing the methods for understanding a user's interests by analyzing those of many other users.

Interestingly, multifaceted recommendation is literally related to the community detection in social networks as described in Sect. 2.6.3, in view of the shared task on identifying strongly-linked users. As such, a straight-forward recommendation approach is suggested for discovering user groups which include the users who are

[1] https://en.wikipedia.org/wiki/Collaborative_filtering.

likely to be interested in the recommended information. For example, based on graph theory, a clustering algorithm [119] was proposed to obtain user/item clusters for recommendation. Saudagar et al. [79] developed a hybrid clustering approach for music recommendation, which uses multimodal information from music profiles and user ratings. Similar clustering based algorithms have also been investigated for the recommendation of articles [71, 163], e-commerce products [104] and scientific publications [137].

Alternately, recommendation can also be addressed from the ranking perspective, which relies on the evaluation of the relevance between users and the recommended items. Matrix factorization (MF) [74, 131, 178] (Similar to NMF as described in Sect. 2.1.5) is a big branch of collaborative filtering, which factorizes the user-item matrix to obtain the latent vectors of both users and items in the same feature space. The similarity between a user and an item is obtained by a product of their latent vectors. More importantly, deep neural networks, as the most effective embedding technique so far, have been incorporated in the MF-based approaches for effective recommendation [73, 86]. Besides the MF approach, recommendation algorithms may also be developed based on search/retrieval and hashing methods [85, 165, 176].

References

1. Ackermann MR, Märtens M, Raupach C, Swierkot K, Lammersen C, Sohler C (2012) Streamkm++: a clustering algorithm for data streams. J Exp Algorithmics (JEA) 17(2.4)
2. Agarwal N, Liu H, Tang L, Yu PS (2012) Modeling blogger influence in a community. Soc Netw Anal Min 2(2):139–162
3. Aichholzer O, Aurenhammer F (1996) Classifying hyperplanes in hypercubes. SIAM J Discret Math 225–232
4. Ailon N, Jaiswal R, Monteleoni C (2009) Streaming k-means approximation. In: Advances in neural information processing systems, pp 10–18
5. Anderson A, Huttenlocher D, Kleinberg J, Leskovec J (2012) Discovering value from community activity on focused question answering sites: a case study of stack overflow. In: KDD, pp 850–858
6. Ankerst M, Breunig MM, Kriegel HP, Sander J (1999) OPTICS: ordering points to identify the clustering structure. In: ACM SIGMOD international conference on management of data, pp 49–60
7. Arthur D, Vassilvitskii S (2007) k-means++: the advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms, pp 1027–1035
8. Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: Membership, growth, and evolution. In: KDD, pp 44–54
9. Bahmani B, Moseley B, Vattani A, Kumar R, Vassilvitskii S (2012) Scalable k-means++. Proc VLDB Endow 5(7):622–633
10. Bandyopadhyay S (2011) Genetic algorithms for clustering and fuzzy clustering. WIREs Data Min Knowl Discov 1(6):524–531
11. Bandyopadhyay S, Saha S (2008) A point symmetry-based clustering technique for automatic evolution of clusters. IEEE Trans Knowl Data Eng 20(11):1441–1457
12. Barbakh W, Fyfe C (2008) Online clustering algorithms. Int J Neural Syst 18(3):185–194
13. Bartfai G (1996) An art-based modular architecture for learning hierarchical clusterings. Neurocomputing 13(1):31–45

14. Becker H, Naaman M, Gravano L (2011) Beyond trending topics: Real-world event identi?cation on twitter. In: Proceedings of international AAAI conference on weblogs and social media, pp 438–441
15. Beitzel SM, Jensen EC, Chowdhury A, Grossman D, Frieder O (2004) Hourly analysis of a very large topically categorized web query log. In: SIGIR, pp 321–328
16. Bekkerman R, Jeon J (2007) Multi-modal clustering for multimedia collections. In: CVPR, pp 1–8
17. Bekkerman R, Sahami M (2006) Semi-supervised clustering using combinatorial mrfs. In: ICML workshop on learning in structured output spaces
18. Bekkerman R, Sahami M, Learned-Miller E (2006) Combinatorial markov random fields. In: ECML, pp 30–41
19. Bekkerman R, Scholz M, Viswanathan K (2009) Improving clustering stability with combinatorial mrfs. In: KDD, pp 99–108
20. Bezdek JC, Hathaway R (2002) VAT: A tool for visual assessment of (cluster) tendency. In: Proceedings of international joint conference neural networks, pp 2225–2230
21. Bickel S, Scheffer T (2004) Multi-view clustering. In: ICDM, pp 19–26
22. Bisson G, Grimal C (2012) Co-clustering of multi-view datasets: a parallelizable approach. In: ICDM, pp 828–833
23. Blooma MJ, Chua AYK, Goh DHL (2011) Quadripartite graph-based clustering of questions. In: International conference on information technology: new generations, pp 591–596
24. Cai D, He X, Li Z, Ma W, Wen J (2004) Hierarchical clustering of www image search results using visual, textual and link information. In: Proceedings of ACM multimedia, pp 952–959
25. Cai X, Nie F, Huang H, Kamangar F (2011) Heterogeneous image feature integration via multi-modal spectral clustering. In: CVPR, pp 1977–1984
26. Caicedo JC, BenAbdallah J, González FA, Nasraoui O (2012) Multimodal representation, indexing, automated annotation and retrieval of image collections via non-negative matrix factorization. Neurocomputing 76(1):50–60
27. Caicedo JC, Moreno JG, Niño EA, González FA (2010) Combining visual features and text data for medical image retrieval using latent semantic kernels. In: Proceedings of the international conference on Multimedia information retrieval, pp 359–366
28. Carpenter GA, Grossberg S (1987) ART 2: Self-organization of stable category recognition codes for analog input patterns. Appl Opt 26(23):4919–4930
29. Carpenter GA, Grossberg S (1990) ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. Neural Netw 3(2):129–152
30. Carpenter GA, Grossberg S (2016) Adaptive resonance theory. Springer, Berlin
31. Carpenter GA, Grossberg S, Markuzon N, Reynolds JH, Rosen DB (1992) Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Trans Neural Netw 3(5):698–713
32. Carpenter GA, Grossberg S, Reynolds J (1991) ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. Neural Netw 4(5):565–588
33. Carpenter GA, Grossberg S, Rosen D (1991) ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. Neural Netw 4:493–504
34. Carpenter GA, Grossberg S, Rosen DB (1991) Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Netw 759–771
35. Cha M, Haddadi H, Benevenuto F, Gummadi KP (2010) Measuring user influence in twitter: The million follower fallacy. In: Proceedings of international AAAI conference on weblogs and social media, pp 10–17
36. Chandrika P, Jawahar C (2010) Multi modal semantic indexing for image retrieval. In: CIVR, pp 342–349
37. Chandrika P, Jawahar CV (2010) Multi modal semantic indexing for image retrieval. In: CIVR, pp 342–349
38. Charikar M, O'Callaghan L, Panigrahy R (2003) Better streaming algorithms for clustering problems. In: Proceedings of the annual ACM symposium on theory of computing, pp 30–39

39. Chaudhuri K, Kakade SM, Livescu K, Sridharan K (2009) Multi-view clustering via canonical correlation analysis. In: ICML, pp 129–136
40. Chen C, Luo J, Parker KJ (1998) Image segmentation via adaptive K-means clustering and knowledge-based morphological operations with biomedical applications. IEEE Trans Image Process 7(12):1673–1683
41. Chen L, Roy A (2009) Event detection from flickr data through wavelet-based spatial analysis. In: CIKM, pp 523–532
42. Chen L, Xu D, Tsang IW, Luo J (2012) Tag-based image retrieval improved by augmented features and group-based refinement. IEEE Trans Multimed (T-MM) 1057–1067
43. Chen Q, Wang G, Tan CL (2012) Web image organization and object discovery by actively creating visual clusters through crowdsourcing. In: Proceedings of international conference on tools with artificial intelligence, 419–427
44. Chen Y, Dong M, Wan W (2007) Image co-clustering with multi-modality features and user feedbacks. In: MM, pp 689–692
45. Chen Y, Rege M, Dong M, Hua J (2007) Incorporating user provided constraints into document clustering. In: ICDM, pp 103–112
46. Chen Y, Tu L (2007) Density-based clustering for real-time stream data. In: Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining, pp 133–142
47. Chen Y, Wang L, Dong M (2010) Non-negative matrix factorization for semisupervised heterogeneous data coclustering. TKDE 22(10):1459–1474
48. De Lathauwer L, De Moor B, Vandewalle J (2000) A multilinear singular value decomposition. SIAM J Matrix Anal Appl 21(4):1253–1278
49. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. J Am Soc Inf Sci 41:391–407
50. Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc, Ser B 39(1):1–38
51. Dhillon IS, Guan Y, Kulis B (2004) Kernel k-means: spectral clustering and normalized cuts. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp 551–556
52. Ding C, Li T, Peng W, Park H (2006) Orthogonal nonnegative matrix tri-factorizations for clustering. In: Proceedings of ACM SIGKDD international conference knowledge discovery and data mining, pp 126–135
53. Ding H, Liu J, Lu H (2008) Hierarchical clustering-based navigation of image search results. In: Proceedings of ACM multimedia, pp 741–744
54. Dong Y, Tang J, Wu S, Tian J, Chawla NV, Rao J, Cao H (2012) Link prediction and recommendation across heterogeneous social networks. In: ICDM, pp 181–190
55. Drost I, Bickel S, Scheffer T (2006) Discovering communities in linked data by multi-view clustering. In: From data and information analysis to knowledge engineering, pp 342–349
56. Escalante HJ, Montes M, Sucar E (2012) Multimodal indexing based on semantic cohesion for image retrieval. Inf Retr 15(1):1–32
57. Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD, pp 226–231
58. Faber V (1994) Clustering and the continuous k-means algorithm. Los Alamos Sci 22:138–144
59. Fasel B, Monay F, Gatica-Perez D (2004) Latent semantic analysis of facial action codes for automatic facial expression recognition. In: Proceedings of international conference on multimedia information retrieval, pp 181–188
60. Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315:972–976
61. Fu Z, Ip HHS, Lu H, Lu Z (2011) Multi-modal constraint propagation for heterogeneous image clustering. In: MM, pp 143–152
62. Fujiwara Y, Irie G, Kitahara T (2011) Fast algorithm for affinity propagation. In: IJCAI, pp 2238–2243

63. Gao B, Liu T, Qin T, Zheng X, Cheng Q, Ma W (2005) Web image clustering by consistent utilization of visual features and surrounding texts. Proc ACM Multimed 112–121
64. Gao B, Liu TY, Zheng X, Cheng QS, Ma WY (2005) Consistent bipartite graph co-partitioning for star-structured high-order heterogeneous data co-clustering. In: Proceedings of international conference on knowledge discovery and data mining, pp 41–50
65. Golub GH, Loan CFV (1996) Matrix computations. Johns Hopkins University Press
66. Gower J, Ross G (1969) Minimum spanning trees and single linkage clustering analysis. J R Stat Soc Ser C 595–616
67. Grossberg S (1980) How does a brain build a cognitive code. Psychol Rev 87(1):1–51
68. Gu Q, Zhou J (2009) Co-clustering on manifolds. In: KDD, pp 359–367
69. Guha S, Meyerson A, Mishra N, Motwani R, O'Callaghan L (2003) Clustering data streams: theory and practice. IEEE Trans Knowl Data Eng 15(3):515–528
70. Gundecha P, Liu H (2012) Mining social media: A brief introduction. Tutor Oper Res
71. Habibi M, Popescu-Belis A (2015) Keyword extraction and clustering for document recommendation in conversations. IEEE/ACM Trans Audio, Speech Lang Process (TASLP) 23(4):746–759
72. Harabagiu S, Lacatusu F (2010) Using topic themes for multi-document summarization. ACM Trans Inf Syst 28(3):1–47
73. He X, Liao L, Zhang H, Nie L, Hu X, Chua TS (2017) Neural collaborative filtering. In: Proceedings of the 26th international conference on world wide web, pp 173–182. International World Wide Web Conferences Steering Committee
74. He X, Zhang H, Kan MY, Chua TS (2016) Fast matrix factorization for online recommendation with implicit feedback. In: Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval, pp 549–558. ACM
75. Hong R, Wang M, Li G, Nie L, Zha ZJ, Chua TS (2012) Multimedia question answering. IEEE Trans MultiMed 19(4):72–78
76. Hsu C, Caverlee J, Khabiri E (2011) Hierarchical comments-based clustering. In: Proceedings ACM SAC, pp 1130–1137
77. Hu X, Sun N, Zhang C, Chua TS (2009) Exploiting internal and external semantics for the clustering of short texts using world knowledge. In: Proceedings of ACM conference on information and knowledge management, pp 919–928
78. Hu X, Tang J, Gao H, Liu H (2013) Unsupervised sentiment analysis with emotional signals. In: WWW, pp 607–618
79. Jadhav SL, Mali MMP (2016) Pre-recommendation clustering and review based approach for collaborative filtering based movie recommendation. Int J Inf Technol Comput Sci (IJITCS) 8(7):72–80
80. Ji X, Xu W (2006) Document clustering with prior knowledge. In: Proceedings of international ACM SIGIR conference research and development in information retrieval, pp 405–412
81. Jiang T, Tan AH (2006) Discovering image-text associations for cross-media web information fusion. In: PKDD, pp 561–568
82. Jiang T, Tan AH (2009) Learning image-text associations. IEEE Trans Knowl Data Eng 21(2):161–177
83. Jing F, Wang C, Yao Y, Zhang L, Ma W (2006) Igroup: web image search results clustering. Proc. ACM Multimed 377–384
84. Joachims T (2003) Transductive learning via spectral graph partitioning. In: ICM, pp 290–297
85. Kalantidis Y, Kennedy L, Nguyen H, Mellina C, Shamma DA (2016) Loh and behold: Web-scale visual search, recommendation and clustering using locally optimized hashing. In: European conference on computer vision, pp 702–718. Springer, Berlin
86. Kim D, Park C, Oh J, Lee S, Yu H (2016) Convolutional matrix factorization for document context-aware recommendation. In: Proceedings of the 10th ACM conference on recommender systems, pp 233–240. ACM
87. Ko J, Si L, Nyberg E (2010) Combining evidence with a probabilistic framework for answer ranking and answer merging in question answering. Inf Process Manag 46(5):541–554

88. Konstas I, Stathopoulos V, Jose JM (2009) On social networks and collaborative recommendation. In: Proceedings of international ACM SIGIR conference on research and development in information retrieval, pp 195–202
89. Kothari R, Pitts D (1999) On finding the number of clusters. Pattern Recognit Lett 20(4):405–416
90. Kriegel HP, Kroger P, Sander J, Zimek A (2011) Density-based clustering. WIREs Data Min Knowl Discov 1(3):231–240
91. Kulis B, Basu S, Dhillon I, Mooney R (2005) Semi-supervised graph clustering: A kernel approach. Proc Int Conf Mach Learn 457–464
92. Kumar A, III HD (2011) A co-training approach for multi-view spectral clustering. In: ICML, pp 393–400
93. Kwok C, Etzioni O, Weld DS (2001) Scaling question answering to the web. ACM Trans Inf Syst (TOIS) 19(3):242–262
94. Lee D, Seung H (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791
95. Lee JS, Olafsson S (2013) A meta-learning approach for determining the number of clusters with consideration of nearest neighbors. Inf Sci 232:208–224
96. Leuken RHV, Garcia L, Olivares X, Zwol RV (2009) Visual diversification of image search results. In: WWW, pp 341–350
97. Leung KWT, Ng W, Lee DL (2008) Personalized concept-based clustering of search engine queries. IEEE Trans Knowl Data Eng 20(11):1505–1518
98. Leung Y, Zhang JS, Xu ZB (2000) Clustering by scale-space filtering. IEEE Trans Pattern Anal Mach Intell 22(12):1394–1410
99. Li M, Xue XB, Zhou ZH (2009) Exploiting multi-modal interactions: A unified framework. In: IJCAI, pp 1120–1125
100. Li M, Xue XB, Zhou ZH (2009) Exploiting multi-modal interactions: A unified framework. In: IJCAI, pp 1120–1125
101. Li MJ, Ng MK, Cheung Y, Huang ZX (2008) Agglomerative fuzzy k-means clustering algorithm with selection of number of clusters. IEEE Trans Knowl Data Eng 20(11):1519–1534
102. Li R, Lei KH, Khadiwala R, Chang KCC (2012) Tedas: A twitter-based event detection and analysis system. Int Conf Data Eng 1273–1276
103. Liang J, Zhao X, Li D, Cao F, Dang C (2012) Determining the number of clusters using information entropy for mixed data. Pattern Recognit 45(6):2251–2265
104. Liao CL, Lee SJ (2016) A clustering based approach to improving the efficiency of collaborative filtering recommendation. Electron Commer Res Appl 18:1–9
105. Liberty E, Sriharsha R, Sviridenko M (2016) An algorithm for online k-means clustering. In: 2016 Proceedings of the eighteenth workshop on algorithm engineering and experiments (ALENEX), pp 81–89. SIAM
106. Lienhart R, Romberg S, Hörster E (2009) Multilayer pLSA for multimodal image retrieval. In: Proceedings of the ACM international conference on image and video retrieval
107. Liu B (2010) Sentiment analysis and subjectivity. Handbook of natural language processing. CRC Press, pp 627–666
108. Liu W, Chang S (2009) Robust multi-class transductive learning with graphs. In: CVPR, pp 381–388
109. Lloyd SP (1982) Least squares quantization in PCM. IEEE Trans Inf Theory 28(2):129–137
110. Long B, Wu X, Zhang Z, Yu PS (2006) Spectral clustering for multi-type relational data. In: ICML, pp 585–592
111. Macropol K, Singh A (2010) Scalable discovery of best clusters on large graphs. In: VLDB Endowment, pp 693–702
112. Massey L (2005) Real-world text clustering with adaptive resonance theory neural networks. In: Proceedings of international joint conference on neural networks, pp 2748–2753
113. McNicholas PD, Murphy TB (2010) Model-based clustering of microarray expression data via latent gaussian mixture models. Bioinformatics 26(21):2705–2712

114. Meng L, Tan AH (2012) Semi-supervised hierarchical clustering for personalized web image organization. In: Proceedings of international joint conference on neural networks (IJCNN), pp 1–8
115. Meng L, Tan AH (2014) Community discovery in social networks via heterogeneous link association and fusion. SIAM Int Conf Data Min (SDM) 803–811
116. Meng L, Tan AH, Wunsch DC (2016) Adaptive scaling of cluster boundaries for large-scale social media data clustering. IEEE Trans Neural Netw Learn Syst 27(12):2656–2669
117. Meng L, Tan AH, Xu D (2014) Semi-supervised heterogeneous fusion for multimedia data co-clustering. IEEE Trans Knowl Data Eng 26(9):2293–2306
118. Messina A, Montagnuolo M (2009) A generalised cross-modal clustering method applied to multimedia news semantic indexing and retrieval. In: WWW, pp 321–330
119. Moradi P, Ahmadian S, Akhlaghian F (2015) An effective trust-based recommendation method using a novel graph clustering algorithm. Phys A: Stat Mech Its Appl 436:462–481
120. Moshtaghi M, Leckie C, Bezdek JC (2016) Online clustering of multivariate time-series. In: Proceedings of the 2016 SIAM international conference on data mining, pp 360–368. SIAM
121. Mumtaz A, Coviello E, Lanckriet GRG, Chan AB (2013) Clustering dynamic textures with the hierarchical EM algorithm for modeling video. IEEE Trans Pattern Anal Mach Intell 35(7):1606–1621
122. Nguyen L, Woon K, Tan AH (2008) A self-organizing neural model for multimedia information fusion. Int Conf Inf Fusion 1–7
123. Nock R, Nielsen F (2006) On weighting clustering. IEEE Trans Pattern Anal Mach Intell 28(8), 1–13
124. Ozsoy MG, Cicekli I, Alpaslan FN (2010) Text summarization of turkish texts using latent semantic analysis. Proc Int Conf Comput Linguist 869–876
125. Paltoglou G, Thelwall M (2012) Twitter, myspace, digg: Unsupervised sentiment analysis in social media. ACM Trans Intell Syst Technol (TIST) 3(4):1–19
126. Papadopoulos S, Kompatsiaris Y, Vakali A, Spyridonos P (2012) Community detection in social media. Data Min Knowl Discov 24(3):515–554
127. Papadopoulos S, Zigkolis C, Kompatsiaris Y, Vakali A (2011) Cluster-based landmark and event detection for tagged photo collections. IEEE Multimed Mag 18(1):52–63
128. Pei T, Jasra A, Hand D, Zhu AX, Zhou C (2009) DECODE: A new method for discovering clusters of different densities in spatial data. Data Min Knowl Discov 18(3):337–369
129. Petkos G, Papadopoulos S, Kompatsiaris Y (2012) Social event detection using multimodal clustering and integrating supervisory signals. In: ICMR, pp 23:1–23:8
130. Pham TT, Maillot NE, Lim JH, Chevallet JP (2007) Latent semantic fusion model for image retrieval and annotation. In: Proceedings of international conference on information and knowledge management, pp 439–444
131. Pirasteh P, Hwang D, Jung JJ (2015) Exploiting matrix factorization to asymmetric user similarities in recommendation systems. Knowl-Based Syst 83:51–57
132. Qin Y, Priebe CE (2013) Maximum lq-likelihood estimation via the expectation-maximization algorithm: A robust estimation of mixture models. J Am Stat Assoc 108(503):914–928
133. Qiu G (2004) Clustering and the continuous k-means algorithm. Proc Int Conf Pattern Recognit 991–994
134. Rasiwasia N, Pereira J (2010) A new approach to cross-modal multimedia retrieval. In: MM, pp 251–260
135. Rege M, Dong M, Fotouhi F (2006) Co-clustering documents and words using bipartite isoperimetric graph partitioning. In: Proceedingd of international conference on data mining, pp 532–541
136. Rege M, Dong M, Hua J (2008) Graph theoretical framework for simultaneously integrating visual and textual features for efficient web image clustering. In: Proceedings of international conference on world wide web, pp 317–326
137. Ren X, Liu J, Yu X, Khandelwal U, Gu Q, Wang L, Han J (2014) Cluscite: Effective citation recommendation by information network-based clustering. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 821–830. ACM

138. Ricci F, Rokach L, Shapira B (2015) Recommender systems: introduction and challenges. In: Recommender systems handbook, pp 1–34. Springer, Berlin (2015)
139. Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. Science 344(6191):1492–1496
140. Sahoo N, Callan J, Krishnan R, Duncan G, Padman R (2006) Incremental hierarchical clustering of text documents. In: Proceedings of ACM international conference on Information and knowledge management, pp 357–366
141. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of international conference on world wide web, pp 851–860
142. Sander J, Ester M, Kriegel HP, Xu X (1998) Density-based clustering in spatial databases: The algorithm gdbscan and its applications. Data Min Knowl Discov 2(2):169–194
143. Satuluri V, Parthasarathy S, Ruan Y (2011) Local graph sparsification for scalable clustering. In: SIGMOD, pp 721–732
144. Schtze H, Silverstein C (1997) Projections for efficient document clustering. In: proceedings of SIGIR, pp 74–81
145. Shental N, Hertz T, Weinshall D, Pavel M (2002) Adjustment learning and relevant component analysis. In: ECCV, pp 776–792
146. Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 22(8):888–905
147. Shi X, Fan W, Yu PS (2010) Efficient semi-supervised spectral co-clustering with constraints. In: ICDM, pp 532–541
148. Silva JA, Faria ER, Barros RC, Hruschka ER, de Carvalho AC, Gama J (2013) Data stream clustering: a survey. ACM Comput Surv (CSUR) 46(1). No. 13
149. Sledge I, Huband J, Bezdek JC (2008) (automatic) cluster count extraction from unlabeled datasets. In: Fifth international conference fuzzy systems and knowledge discovery (FSKD), pp 3–13
150. Sriram B, Fuhry D, Demir E, Ferhatosmanoglu H, Demirbas M (2010) Short text classification in twitter to improve information filtering. In: International ACM SIGIR conference on research and development in information retrieval, pp 841–842
151. Steinberger J, Jezek K (2004) Using latent semantic analysis in text summarization and summary evaluation. In: Proceedings of ISIM, pp 93–100
152. Su JH, Wang BW, Hsu TY, Chou CL, Tseng VS (2010) Multi-modal image retrieval by integrating web image annotation, concept matching and fuzzy ranking techniques. Int J Fuzzy Syst 12(2):136–149
153. Sugar CA, James GM (2003) Finding the number of clusters in a data set: an information theoretic approach. J Am Stat Assoc 98(463):750–763
154. Sun H, Wang S, Jiang Q (2004) FCM-based model selection algorithms for determining the number of clusters. Pattern Recognit 37(10):2027–2037
155. Tan AH, Carpenter GA, Grossberg S (2007) Intelligence through interaction: Towards a unified theory for learning. LNCS 4491:1094–1103
156. Tan AH, Ong HL, Pan H, Ng J, Li Q (2004) Towards personalised web intelligence. Knowl Inf Syst 6(5):595–616
157. Tang L, Wang X, Liu H (2009) Uncovering groups via heterogeneous interaction analysis. In: ICDM, pp 503–512
158. Tang W, Lu Z, Dhillon IS (2009) Clustering with multiple graphs. In: ICDM, pp 1016–1021
159. Tran TN, Wehrens R, Buydens LMC (2006) KNN-kernel density-based clustering for high-dimensional multivariate data. Comput Stat Data Anal 51(2):513–525
160. Wang D, Zhu S, Li T, Chi Y, Gong Y (2011) Integrating document clustering and multidocument summarization. ACM Trans Knowl Discov Data 5(3):1–26
161. Wang L, Leckie C, Ramamohanarao K, Bezdek J (2012) Automatically determining the number of clusters in unlabeled data sets. IEEE Trans Knowl Data Eng 21(3):335–350
162. Wang W, Zhang Y (2007) On fuzzy cluster validity indices. Fuzzy Sets Syst 158(19):2095–2117

163. West JD, Wesley-Smith I, Bergstrom CT (2016) A recommendation system based on hierarchical clustering of an article-level citation network. IEEE Trans Big Data 2(2):113–123
164. Whang JJ, Sui X, Sun Y, Dhillon IS (2012) Scalable and memory-efficient clustering of large-scale social networks. In: ICDM, pp 705–714
165. Wu B, Zhou X, Jin Q (2015) Participatory information search and recommendation based on social roles and networks. Multimed Tools Appl 74(14):5173–5188
166. Wu Z, Leahy R (1993) An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. IEEE Trans Pattern Anal Mach Intell 15(11):1101–1113
167. Xing E, Ng A, Jordan M, Russell S (2003) Distance metric learning with application to clustering with side-information. In: NIPS, pp 505–512
168. Xu R, Wunsch D (2005) Survey of clustering algorithms. IEEE Trans Neural Netw 16(3):645–678
169. Xu R, Wunsch D (2008) Clustering, vol 10. Wiley, New York
170. Xu W, Liu X, Gong Y (2003) Document clustering based on non-negative matrix factorization. In: Proceedings of SIGIR conference on Research and development in informaion retrieval, pp 268–273
171. Jia JY, Wang CZ, Hua XS (2008) Finding image exemplars using fast sparse affinity propagation. In: ACM MM, pp 639–642
172. Yan H, Chen KK, Liu L, Bae J (2009) Determining the best k for clustering transactional datasets: A coverage density-based approach. Data Knowl Eng 68(1):28–48
173. Yang J, Leskovec J (2012) Defining and evaluating network communities based on ground-truth. In: SDM, pp 745–754
174. Yang Y, Chawla N, Sun Y, Han J (2012) Predicting links in multi-relational and heterogeneous networks. In: ICDM, pp 755–764
175. Yin J, Wang J (2016) A text clustering algorithm using an online clustering scheme for initialization. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1995–2004. ACM
176. Zamani H, Croft WB (2018) Joint modeling and optimization of search and recommendation. arXiv:1807.05631
177. Zhang K, Lo D, Lim EP, Prasetyo PK (2013) Mining indirect antagonistic communities from social interactions. Knowl Inf Syst 35(3):553–583
178. Zhang Y, Chen M, Huang D, Wu D, Li Y (2017) idoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. Futur Gener Comput Syst 66:30–35
179. Zhang Z, Nasraoui O (2006) Mining search engine query logs for query recommendation. In: WWW, pp 1039–1040
180. Zhao R, Grosky W (2002) Narrowing the semantic gap improved text-based web document retrieval using visual features. IEEE Trans Multimed pp 189–200
181. Zhou D, Burges CJC (2007) Spectral clustering and transductive learning with multiple views. In: ICM, pp 1159–1166
182. Zhu L, Galstyan A, Cheng J, Lerman K (2014) Tripartite graph clustering for dynamic sentiment analysis on social media. In: SIGMOD, pp 1531–1542