

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337939689>

Survey on Social Networks Data Analysis

Chapter · December 2019

DOI: 10.1007/978-3-030-37484-6_6

CITATIONS

4

READS

1,040

4 authors, including:



Soufiene Jaffali

Higher Institute of Technological Studies of Kebili

9 PUBLICATIONS 43 CITATIONS

[SEE PROFILE](#)



Salma Jamoussi

University of Sfax

119 PUBLICATIONS 605 CITATIONS

[SEE PROFILE](#)



Abdelmajid Ben Hamadou

University of Sfax

260 PUBLICATIONS 1,434 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



A new committee based method for Bayesian Network structure learning: Application to gene regulatory networks [View project](#)



Formal development of lingware systems [View project](#)

Survey on Social Networks Data Analysis

Soufien Jaffali¹[0000–0003–4612–0991], Salma Jamoussi²[0000–0003–1521–3351],
Nesrine Khelifi¹[0000–0002–7466–151X], and Abdelmajid Ben
Hamadou²[0000–0001–8632–2956]

¹ Imam Abdulrahman Bin Faisal University, Dammam 34212, Saudi Arabia
{smjaffali, nmkhelifi}@iau.edu.sa.com

² Sfax University, Sfax 3029, Tunisia
{jamoussi, abdelmajid.benhamadou}@gmail.com

Abstract. Social networks are the most successful Web 2.0 applications, where users share and create over 2.5 quintillion bytes of data daily. This data can be exploited to retrieve many kinds of information which will be used in several applications. In fact, social networks have attracted considerable attention from researchers in different domains. This paper serves as an introduction to social network data analysis. In this work we present the recent and representative works in social network data analysis in an analytical fashion. We also highlight most important applications and used methods in the context of structural data analysis. Then, we list the major tasks and approaches proposed to analyse added-content in social media.

Keywords: Social network · Data analysis · Structural analysis · Added-content analysis.

1 Introduction

Recently, there has been a growing interest in social networks, given their impact on society, economy and many sides of user lives. One of the most important characteristics of social networks is the huge amount of data. This data represents a precious mine of information which attracted researchers from different domains. In this context, most of literature works focused on the social network topology or the communications between individuals [24], which represents the linkage information. Several purposes motivated researchers to analyze link information in social networks such as studying the network evolution [84], predicting future links [47] and discovering communities within the social network [19]. Other works analyze the added-content to social networks such as text and multimedia. This content has an important added-value to the analysis.

Previous surveys studying social network data analysis focused on a single analysis axis such as community detection [19] or sentiment analysis [41]. Few surveys tried to present the state of the art of social network data analysis in general without focusing on a special field [2]. The literature review in these surveys organizes works according to the application like community discovery

and node classification. [69] presents a survey on social networking data analysis frameworks classified according the kind of analysis.

In this paper, a new social network data analysis survey is proposed, in which we organize works according to the nature of the analysed data. Thus, we cluster the most representative works into structured-based analysis and added-content based analysis. The remainder of the paper is organized into three sections. Section 2 outlines the most representative works focused on social network structural data analysis. Section 3 describes methods and approaches used to analyze the added-content. Section 4 concludes the paper.

2 Structural Analysis

The basic idea of the structural analysis approaches is that the users are rather characterized by their relations than by their attributes (name, age, social class, etc.) or their added-content (what they share) [57].

The density of relationships and the distances or the number of hops between users within the social network are variable. Furthermore, some actors occupy more central positions than others.

The purpose of structural analysis is therefore to study these phenomena, to identify strong links and weak links, structural holes and central nodes. In this context, some researchers studied links within the social network to extract nodes with higher impact and to determine communities and regions in evolution. Among the major research scopes relying on structural analysis, we can cite link prediction, community detection, user classification, social network evolution analysis, the link inference and visualization.

2.1 Community Detection

There is no universal or exact definition of a “community”. Communities are also called groups, clusters, coherent subgroups, or modules in different contexts. In sociology, a community is formed by individuals interacting with each other more often than those outside the community. Community detection and tracking [19] can facilitate other social analytic tasks and its results are exploited in many real-world applications. For example, bringing together customers with similar interests in social media can lead to more effective recommendations that expose customers to a wide range of relevant articles. Communities can also be used to compress a huge network, by treating the set of users forming a community as a single node. In other words, problem-solving is done at the group level, instead of the user level. In the same vein, an extensive network can be viewed at different resolutions, providing an intuitive solution for network analysis and visualization.

Some methods create communities in social networks based on similarities or distances between nodes (users) within the network [53]. Other works retrieve communities by dividing the social network into a set of high densely interconnected nodes [28]. The majority of community detection methods deals with

static networks [24], ignoring the time dimension in social networks. To overcome this drawback, some new methods are proposed to take into consideration the dynamic aspect [68].

2.2 User Classification

User classification is a task closely related to community detection. The huge amount of information collected about users, their connections, opinions, activities and thoughts can be modeled as labels associated with individuals. Several forms of labels are possible such as demographic labels (age, gender and place), political labels, labels that encode religious beliefs, labels that represent interests, preferences, and affiliations.

Unfortunately, this information is not available about all users. Some of users are either partially labeled or unlabeled. The classification task aims to infer the unknown user labels based on correlated labeled users. In literature, several models and methods are proposed for user classification such as Relational Neighbor classifier [52], latent group model [59] and Copuler Latent Markov model [83].

2.3 Social Network Evolution Analysis

The dynamism is one of the most important aspects of social networks, which is a manifestation of interactions between individuals. Indeed, the structure of the network can evolve over time thanks to the change of roles, social status and even user preferences. There are a variety of approaches and algorithms available to analyze the evolution of social networks. Guandong et al. [84] classify these algorithms into two families.

The first family uses an additional temporal dimension with the traditional graphic representation (two-dimensional graph, where the two dimensions of the graph can be "*author* \times *keyword*" or "*user* \times *user*") [7, 74]. In this approach, the network in the different stamps is modeled as a two-dimensional graph, represented by a matrix. The new representation is thus in the form of a three-dimensional cube of each data slice corresponding to the matrix representing the network in a specific stamp.

The second category consists in simultaneously discovering the structure of communities and the dynamic network evolution. Unlike classic methods of network evolution analysis and community detection, this category takes into account both network evolution and community detection to analyze the social network. These approaches are based on the assumption that the evolution of community structures in a social network from one stage to its next will not be a dramatic change. In [48], Lin et al. provide a general framework for the analysis of communities and their evolution in dynamic networks. In their approach, the authors measure the quality of a community structure at a certain time t using a cost function *Cost*. The formation of a stable community structure depends on the minimization of the cost function. The latter is composed of an instantaneous cost *CI* and a time cost *CT*.

$$Cost = \alpha \cdot CI + (1 - \alpha) \cdot CT \quad (1)$$

The instantaneous cost CI is then determined by the difference in distribution between the similarity matrix of the nodes in the network and the calculated similarity matrix of the community structures formed. In other words, the instantaneous cost is the Kullback-Leibler divergence between the aforementioned similarity matrices. On the other hand, the time cost CT indicates the difference in distribution between the similarity matrices at two consecutive time stamps using the same formula. Thus, the analysis of community structures and their evolution is converted into an optimization problem to find appropriate communities that minimize the total cost.

2.4 Link Prediction

As we have already explained, social networks are dynamic and in continuous evolution. Indeed, it is always possible to add nodes and links in the network. Understanding evolution and, above all, being able to predict the next links that will be established in social networks is a very active topic of research, known as link prediction. According to [47], the prediction techniques of links in social networks can be categorized into two main approaches. The first approach includes proximity-based methods, which use node information, topology, or social theory to compute the similarity between a pair of nodes. In the second approach, we find learning-based methods of predicting links based on machine learning.

Proximity-based methods

Nodal proximity-Based method: Intuitively it's the simplest idea: the probability of having link in the future between two nodes is proportional to the similarity between them [6]. This idea is consistent with the fact that a person tends to establish relationships with people who resemble him in religion, location or education.

A node in a social network is usually represented by its name, publications, interests, or demographic information. It is therefore sufficient to use these attributes to measure the similarity between a pair of nodes. The greater the similarity between an unconnected pair of nodes, the greater the likelihood that they will be linked. On the other hand, a low similarity indicates that it is very likely that the pair of nodes will not be bound in the future.

In the work of Anderson et al. [5], the similarity between the nodes is measured using the overlap between the users' interests. User interests are extracted from their actions such as editing an article on *Wikipedia* or asking a question about *Stack Overflow*. The interests of each user are represented as a vector, and the similarity between two users is calculated using the cosine as a measure of difference between their vectors of interest.

Topology Based method: In addition to the characteristics of nodes in social networks, the network topology can also be used to determine the similarity between two nodes, and predict the possibility of linking them. Depending on their characteristics, topology-based link prediction method can be categorized into neighbor-based method, path-based method, and random walk method.

- *Neighbor-based methods:* Among these methods we can cite *Common Neighbors (CN)*. This method, known for its simplicity, is widely used to predict links [60]. His principle is simple; to measure the similarity between two nodes x and y , it is enough to calculate the number of nodes in direct interaction with the nodes x and y at a time. The *CN* measure is defined as follows:

$$CN(x, y) = |\Gamma(x) \cap \Gamma(y)| \quad (2)$$

With $\Gamma(x)$ (respectively $\Gamma(y)$) is the set of neighboring nodes of the x node (respectively y), and $|\Gamma(x)|$ is the number of nodes in the $\Gamma(x)$ set. The higher the value of *CN*, the more the x and y nodes have a chance to be linked. Since the *CN* measure is not normalized, it reflects a relative similarity between the pairs of nodes. Several methods are proposed to normalize it. We cite as an example the measure of *Leicht-Holme-Nerman (LHN)* [45], defined as follows:

$$LHN(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x)| \cdot |\Gamma(y)|} \quad (3)$$

Thus, we assign a higher similarity to nodes having a large number of neighbors in common compared to the total number of their neighbors.

- *Path-Based Methods:* These methods consist of calculating the number of paths between the x and y nodes to measure the similarity between them. Lü et al. [51] use the *Local Path (LP)*, to measure the similarity between a pair of nodes. This measure consists of calculating the number of paths of length 2 and length 3 between the two nodes. Given that paths of length 2 are more relevant, a factor of α (small value close to 0) is applied to adjust the importance given to paths of length equal 3.

$$LP(x, y) = A^2 + \alpha A^3 \quad (4)$$

Where A^i is the adjacency matrix of the nodes separated by paths of length i . The *FrienLink (FL)* method proposed in [62] provides a faster and more efficient prediction. It follows the principle that two people in a social network can use all the paths between them to establish a direct link proportionally to the lengths of these paths. *FL* is defined by the following function:

$$FL(x, y) = \sum_{i=1}^l \frac{1}{i-1} \cdot \frac{|path_{x,y}^i|}{\prod_{j=2}^i (n-j)} \quad (5)$$

With l is the path length between x and y (excluding paths containing cycles), n is the number of vertices in the network, $path_{x,y}^i$ is the set of all paths of length i between x and y .

- *Random walk-based methods*: The random walk technique consists of calculating a probability for each neighbor of the node x , to determine to which node a random walker located at x will move. Some methods are based on the random walk to measure the similarity between a pair of nodes. Among these methods *Time of Strike (TS)* [25] calculates the expected number of steps a random walker do to reach the node y starting from x . This measure is given by the following function:

$$TS(x, y) = 1 + \sum_{w \in \Gamma(x)} P_{x,w} TS(w, y) \quad (6)$$

With $P = D_A^{-1}A$, where D_A is the diagonal matrix of A .

Social Theories Based Metrics: Some recent works like [50, 77] introduce social theories such as *triadic closure* and *homophily* to refine the quality of link prediction. In [77], Valverde et al. combine topology-based methods (*Common Neighbors*) with user behavior and interest information to predict future links in *Twitter*.

Liu et al. [50] propose another model of link prediction using the notion of node centrality³. In this model, neighboring nodes participate in the establishment of new links in proportion to their centrality. The introduced module is given by:

$$S(x, y) = \sum_z (w(z) \cdot f(z))^\beta, \text{ with } f(z) = \begin{cases} 1 & \text{if } z \in \Gamma(x) \cap \Gamma(y) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Where $w(z)$ presents the centrality weight of the node z , and β is a constant between -1 and 1 used to adjust the contribution of each common neighbor to the probability of linking the nodes x and y .

Learning-Based Link Prediction Some recent works propose models based on machine learning for the prediction of links [14, 46, 54]. These models can be classified into three categories, namely: classification according to characteristics, probabilistic graph models and matrix factorization.

Characteristic-Based Classification: Let $G(N, E)$ be a graph corresponding to a social network, with N being the set of nodes and E being the set of edges between N . $l^{(x,y)}$ is the label corresponding to the instance of nodes (x, y) . Each pair of nodes corresponds to an instance that includes a tag l and a set features describing the pair of nodes. If there is a link between the nodes x and y , the label $l^{(x,y)}$ is a positive value, otherwise it is negative.

$$l^{(x,y)} = \begin{cases} +1 & \text{if } (x, y) \in E \\ -1 & \text{if } (x, y) \notin E \end{cases} \quad (8)$$

³ A measure to capture the vague notion of importance in a graph, used to identify the most significant vertices.

To classify the instance labels, several classification methods can be used, such as *SVM* and *decision trees*. To adopt the classification results to predict links, it is necessary to define and extract an appropriate set of characteristics from the social network. These characteristics are given by methods based on the topology and social theories.

In this context, [46] present a machine-learning model based on a graph kernel and properties of nodes such as age, grade level, and so on. This model is used to predict the links between users and items in a bipartite network.

In the work of Wu et al. [82], the authors propose a three-step model for predicting links: (i) choose candidate nodes using methods based on social theory (*homophilia*). (ii) refine the list of candidates using a factor graph template to adjust the ranking of candidate nodes. (iii) Use the interactive learning method *RankFG+*, which uses users' feedback to effectively update the existing prediction model.

Probabilistic Graph Models: In this type of method, a social network can be represented with a graph, in which one assigns probability values (such as the probability of transition of a random walker, the topological similarity, etc.). The probabilistic graph obtained is widely used for predicting links.

Clauset et al. [14] propose a model of probabilistic graph to infer the hierarchical structure of a network, and apply it for link prediction. Let H be a dendrogram corresponding to the hierarchical structure of the G probability graph. Each leaf of the dendrogram corresponds to a node of the graph and each internal node of the dendrogram corresponds to a probability p_r (the probability of connecting a pair of nodes having r as ancestor).

Let E_r be the number of edges in H , where the closest common ancestor is r and let R_r (respectively L_r) the number of leaves in the right sub-tree (respectively left) rooted at r . Then, the probability of the hierarchical graph is given by:

$$L(H, \{p_r\}) = \prod_r p_r^{E_r} (1 - p_r)^{R_r L_r - E_r} \quad (9)$$

For a given graph G , there are several possible dendrograms. Assuming that all hierarchical random graphs are priory equiprobable, the likelihood that a given model $(H, \{p_r\})$ is the right explanation of the data in G is proportional to the likelihood $L(H, \{p_r\})$ with which this model generates the observed network. To predict whether two unconnected nodes x and y will be connected in the future, we consider a set of dendrogram samples, with probability proportional to their likelihood. Then, we calculate the average of the p_{xy} probabilities on the dendrogram samples.

Matrix Factorization Based Models: Menon et al. [54] treat the prediction of links as a matrix completion problem and extend a method of factorizing the matrix to solve the problem of predicting links. They factorize the graph $G \approx L(U\Lambda U^T)$: With $U \in \mathbb{R}^{n \times k}$, $\Lambda \in \mathbb{R}^{k \times k}$ and $L(\cdot)$ a link function (n is the number of nodes and k is the number of latent characteristics).

Each node x has a latent vector corresponding to $u_x \in \mathbb{R}^k$. Thus, the predicted score for the pair of nodes (x, y) is $L(u_x^T \Lambda u_y)$. This model combines the latent characteristics with explicit characteristics of the nodes and the links in the graph via a bilinear regression model.

2.5 Social Network Visualization

Social network visualization may be an alternative solution to better understand social networks if a list of statistics does not offer a clear vision. In this context, several representations have been proposed where graphic elements represent the individuals and lines to symbolise the links between individuals. We can distinguish four classes of social media visualization methods depending on the purpose of the visualization and the knowledge to be identified.

Structural Visualization The structural visualization of the social network focuses on the network structure, which can be considered as the graph topology that represents individuals and links in the social network. The two main structural visualization approaches are the *node-link diagram* and the *oriented matrix* method.

The first approach is to represent the actors by nodes taking geometric shapes (circles, triangles, etc.) and the relationships between the actors by links between these geometric shapes. In the *diagram node-link*, different properties of the social network such as the importance of a node and the strength of the friendship bond can be expressed via several geometric properties such as color, size, shape, or thickness [26].

The *oriented matrix* method consists of an explicit display of the adjacency or incidence matrix of a social network. In this representation, each link is in the form of a location on a grid having the Cartesian coordinates corresponding to the nodes of the network concerned by the link. Colors and opacity are often used in this kind of visualization to represent important structural or semantic quantities [23].

Semantic Visualization It is a visualization that uses ontologies to represent the types of actors and relationships in the network. The purpose of this representation is the visualization of the attributes of nodes and links.

Ontovis [72] is an example of ontology-based visualization, where a *diagram node-link* is enriched by an ontology graphic representation.

Temporal Visualization Time is an important semantic information that has attracted the researchers' attention. Indeed, individual's behavior and relations are time-dependent. Thus, the time dimension should be visualized to enhance the comprehension. In this context, several types of visualizations are proposed.

Moody et al. [56] propose two types of dynamic visualization, namely; *flip-books* and *movies*. In the first representation, the nodes remain static and relationships change over time. In the second representation, the nodes may vary as the relationships change.

In *Movibis* [71] an ontology-based visualization is enriched by a time graph, which includes spatiotemporal information about the actors in the network.

Statistic Visualization: Another aspect that seems interesting for the visualization of social networks is the statistical aspect. Social network statistics represent the degree, centrality, and coefficient of clustering in networks or communities of users.

One way to understand these statistical distributions is via visual summaries, such as histograms that show the variation of a variable. An additional insight can be obtained by analyzing the joint distribution between two variables at once. This overview is usually obtained through scatter plots, where the points are mapped into a Cartesian coordinate system based on the values of the two variables in question.

In the context of social networks, scatterplots are useful, especially in the representation of edge correlations. This technique has been explored by Kahng et al. [29] to study the correlation of intermediate centrality in a social network.

3 Added Content-Based Analysis

The amount of user-added information in certain networks, such as "Youtube", "Flickr"⁴ and email networks, is gigantic. Therefore, the exploitation of this information is very important to improve the quality of the analysis. In addition, in some networks, more knowledge can be derived from the added-content than through links between nodes. In addition, by analyzing the content of social networks we can better understand the opinions of users on a given subject, identify groups among the masses of users and detect influential people.

Research axis on content-based analysis can be classified according to the nature of the analyzed data. Thus we find text mining in social networks, multimedia mining and sensors and flows mining.

3.1 Text Mining in Social Networks

Social networks are rich in text as they offer users a variety of ways to contribute with text. *Facebook*⁵, for example, allows users to create different textual content, such as publications on the wall, comments on publications and links to other web pages. This fact encouraged researchers to invest in modeling text mining tools to analyze social networks content. In the literature, a variety of algorithms are developed in the context of text mining. However, social networks present

⁴ <https://www.flickr.com/>

⁵ <https://www.facebook.com/>

new challenges because the network structure intervenes to add information to search. In the remainder of this paper, we designate by textual publications all types of textual contributions possible on social networks (tweets, comments, statutes).

Keyword Research In a keyword research, we specify a set of terms to identify the nodes in a social network that are relevant to a given query. The main challenge for keyword research in social networks is how to effectively explore the social network and find the sub-network that contains all the keywords in the query. To cope with this challenge, we use the content and the links together to improve the search result, following the assumption that "documents containing similar terms are often linked" [2].

We can classify keyword research algorithms in social networks into two categories. The algorithms of the first category search the sub-network corresponding to the keywords by exploring all the links of the network, without exploiting any index [39]. The algorithms in the second category are based on a network index, which is used to guide the exploration of the network [32].

Classification In the classification problem, the nodes in the social network are associated with classes. These tagged nodes are subsequently used for learning. There is a whole range of textual content classification algorithms, but the presence of links often provides useful help for classification.

In the case of social networks, there are some of additional challenges to text classification algorithms. These challenges are:

- Social networks contain a much broader and non-standard vocabulary, compared to conventional text collections.
- Tags in social networks are often very rare. Thus, data from social networks are often much more sparse than in other standard text collections.
- The presence of links between nodes can be useful to guide the classification process.

Irfan et al. [34] distinguish three families of text classification methods in social networks:

Ontology-Based Methods: Ontologies can be used for text classification [85], in order to introduce an explicit specification of concepts and their semantic relations [22]. However, semantic analysis is very expensive and difficult in the context of classification of large body of text, as is the case of social networks online [85].

Machine Learning-Based Methods: In this family, we can cite the most used supervised learning algorithms in the text classification context, such as *Rocchio's algorithm* [67], *learning-based 'instances'* [11], *artificial neural networks* [35, 37] and *genetic algorithms* [43].

Hybrid Methods: In the literature, some studies show that the combination of different classification algorithms gives better results and improves the performance of text classification compared to the application of a unique classification method [1, 55]. However, the result of applying a hybrid approach depends largely on test data sets. Therefore, there is no guarantee that we will achieve the same performance by changing the test set [34].

Clustering The clustering problem arises quite often in the context of node grouping. It is closely related to the traditional problem of graph partitioning. Some recent works use text content in social network nodes to improve the quality of grouping [30, 63]. The problem of categorization has been widely studied by the research community in text mining. In this context, a variety of algorithms have been proposed [3]. These algorithms use several variants of traditional categorization algorithms for multidimensional data. Most of these algorithms are variants of the *K-means* method [2].

According to the study by Irfan et al. [34], text categorization methods can be grouped into three large families:

Hierarchical Clustering: Hierarchical categorization organizes a group of documents in a tree structure (dendrogram) where parent/child relationships can be considered topic/subtopic relationships [42].

Partitional Clustering: Partitional clustering groups probabilistic *partitioning* methods such as the *maximization of hope* [8], the *single-pass* [66], *K-means* variations such as *C-means* [12], *k-medoid* [36] and *C-medoid* [86].

Semantic Clustering: Consists of using an external semantic source (*WordNet*, *Wikipedia*), if it exists, or creating a source and using it to find the correlation between the terms. In the context of categorizing short text (such as mini-blogs like *Twitter*), limited knowledge is available, hence it is important to create semantic relationships based on internal semantics [70].

Topic Detection Topic detection and tracking is one of the emerging subject in social network data analysis. It consists on discovering the latent patterns or structures of textual content within social networks. Several methods are proposed to discover topics in the literature. *Latent Dirichlet Allocation* and *Author-Topic Model* are widely used to identify topics in social network [33]. Authors in [15] proves that SVMs are efficient to predict users political alignment based on *Twitter* hashtags.

Learning Transfer Many social networks contain heterogeneous content, such as text, video, and images. Some types of content may have more or less difficulty in the process of learning classification or categorization. Indeed, several forms of data can not be available to the user in certain fields. For example,

it is relatively easy to obtain learning data for textual content. However, this is not always true in the case of images, where there are fewer collections for learning. In this case, the learning transfer is useful for solving classification or grouping problems caused by the lack of data availability. Also, learning transfer is often useful for transferring knowledge from one form of data to another by using the concept of *mapping*. A similar observation applies to the case of text classification problems in cross languages. Indeed, the amounts of learning data available for some languages are not always sufficient. For example, a large amount of labeled textual content may be available in the English language; which is not true in some other languages such as Arabic [44]. Pan and Yang [61] distinguish three categories of techniques used for learning transfer, namely: *Inductive Learning Transfer* [65], *Transductive Learning Transfer* [20, 88] and *Unsupervised Learning Transfer* [18, 79].

3.2 Multimedia Data Mining

A multimedia information network is a structured collection, where multimedia documents, such as images and videos, are conceptually represented by nicknames tied by links. These links correspond either to real hyperlinks in social networks and web pages, or to logical relationships between the nodes that are implicitly defined through meta-information such as user identifiers, location information or other labels.

Multimedia news networks are seen as a marriage between multimedia content and social networks. They are richer in information thanks to the information stored in the interaction between these two types of entities. To understand multimedia information networks, we must not consider the visual characteristics of each node, but also explore the network structure with which they are associated. Recently, several studies have been carried out in this field, but which have not yet reached the desired maturity. It is important to note that the link structure in multimedia information networks is essentially logical. These links are created from logical common points in the characteristics and relationships between the different entities of the network. Aggarwal [2] distinguishes four sources of logical links that can be used to create a multimedia information network besides the hypertext links often present in social networks.

Ontologies In multimedia information networks created from ontologies, some studies have focused on the hierarchical classification of concepts. Kamvar et al. [40] associates each node of the network with a binary classifier that calculates the conditional probability given the previous node. the network is constructed by exploiting the semantic relationship "is-a" and "part of" in a general graph, since the graph may not be a tree, there may be several paths from the node (concept) root to any target concept. Each path is associated with the minimum conditional probabilities associated with all edges of the trajectory. The marginal probability of any target concept is defined as the maximum of all likely paths connecting the root concept to the target concept. The same classifier is used

for the classification of concepts in the network *ImageNet* [21]. Another method of more sophisticated hierarchical classification is proposed by Cai and Hoffman [10]. This method consists in encoding the root-to-target concept path through a fixed-size binary vector, and treat it as a multi-label classification problem using SVM.

Media Communities In media communities, users play a central role in multimedia content indexing and retrieval. The basic idea is quite different from a multimedia system centered on traditional content. Web sites that provide media communities are not only run by site owners, but by millions of amateur users who provide, share, edit, and index content.

In this context, some works have focused on the development of multimedia search systems by exploiting the tags provided by users [64, 78]. As a result, two basic problems must be solved. First, user tags are often very noisy or semantically meaningless [13]. More specifically, user tags are known to be ambiguous, limited in terms of completeness, and too personalized [2]. Secondly, the tags associated with an image, for example, are generally in random order without any information of importance or relevance, which limits the effectiveness of these tags in the search and other applications.

To remedy the first problem, Tang et al. [76] suggest building an intermediate space of concepts from user tags. This space can be used as a means of deducing and detecting new concepts more generic in the future. The work in [80] provides a probabilistic framework for resolving ambiguous tags that are likely to occur, and which appear in different contexts. There are also many tag suggestion methods that allow users to annotate media with more informative labels, and avoid low-quality or meaningless labels [4, 73].

To overcome the second problem, Liu et al. [49] propose a tag classification system that aims to automatically classify the tags associated with a given image according to their relevance to the image content. This system estimates the initial tags relevance scores based on an estimations of likelihood density, followed by a random walk on a tag similarity graph to refine the relevance scores.

Personal Photo Albums Among the major lines of research that are interested in personal photo albums in social networks, we can mention the discovery of relationships between people in the same photos. Wu et al [81] took advantage of face categorization technology to discover the social relationships of subjects in personal photo collections. The co-occurrence of identified faces as well as the distances between the faces in the personal photos are used to calculate the strength of the connection between two identified persons. Also, social classes and the social importance of individuals can be calculated.

The labeling and categorization of personal photo albums in social networks is another attractive area of research. In this context, some works are interested in the labeling and categorization of personal photos by exploiting the information on the date and place of taking photos [16, 75]. The motivation behind this work

is that events and location are among the best clues that people remember and use when searching for photos [58].

Geographical Information Thanks to the progress of low-cost GPS (Global Positioning System ⁶), cell phones and cameras have become equipped with GPS receivers, and so are able to save locations by taking pictures. As a result, the use of geographic information has become increasingly popular. Furthermore, social networks allow users to specify the location of their shared pictures. As a result, millions of new geo-tagged images are added each month to *Flickr*.

Geographical annotation is a rich source of information that can be used in many applications. Among these applications, we find the use of geographic information for a better semantic understanding of the image. Joshi and Luo [38] propose to explore databases of geographic information systems (*SIG* ⁷) using a given geographical location. They use descriptions of small local neighborhoods to form *geo-tags* bags. The association of *geo-tags* with visual features is used as a basis for learning to derive the label of an event or activity such as "natural disaster" or "marriage". The authors demonstrate that the context of geographic location is a good indicator for recognition of an event. Yu and Luo [87] improves the accuracy of region recognition by using a probabilistic graphical model. At the entrance of the model, they provide the information of the merged visual and non-visual context. Information on location and time is obtained through image metadata automatically.

Another line of research is devoted to the estimation and deduction of geographic information. The work in this area is motivated by the fact that many web users may be interested in where a beautiful photo is taken and want to know where exactly it is. Also, if a user plans to visit a place, he may want to discover places of interest nearby. In this context, Hays and Efros [31] collected millions of geolocated *Flickr* images. Using a comprehensive set of visual features, they used nearest neighbor search in the reference set to locate the image. Motivated by the work of Hays and Efros [31], Gallagher et al. [27] incorporate text tags to estimate the geographic locations of images. Their results show that text tags perform better than visual content, but they do better by combining visual content with text tags. Crandall et al. [17] estimate the approximate location of a new photo using an SVM classifier. The new image is geolocated by assigning it to the best cluster based on its visual content and annotations.

3.3 Sensors and Data Flows Mining

Driven by the explosive development of online social networks such as "*Facebook*", "*Twitter*", and the growing popularity of smartphones, and even the rapid evolution of sensor networks, mobile social networks are likely to become the future key direction for mobile computing. Beach et al. [9] defines mobile social networks as a distributed system that combines the three components: telephone,

⁶ https://en.wikipedia.org/wiki/Global_Positioning_System

⁷ https://en.wikipedia.org/wiki/Geographic_information_system

social and sensor. Telephone devices are ubiquitous and can provide information on the identity or location of the user. In addition, smartphones are equipped with a variety of sensors such as accelerometers, microphones, cameras and even digital compasses, which can be leveraged to infer user orientations and preferences. However, telephone devices alone can not produce a complete picture of the context, especially user preferences. On the other hand, social networks are rich in detailed contextual information describing an individual's personal interests and relationships, but they can not locate users. Sensor networks provide the desired third dimension of the data and thus provide a better understanding of the context. Taking the example of information provided by sensors such as temperature, humidity, sound and video and everything characterizing the individual's context. In addition, data from the three classes: telephones, social, and sensors when archived together can improve the context understanding. Therefore, it is the combination of these three data streams that allows the consideration of the current location and the current preferences of the user to be able to perform conscious actions of the user's context, such as movie recommendation or friend suggestion.

4 Conclusion

In this paper, we presented a state of the art of social network analysis methods. We organized these methods, according to the type of analyzed data, in two main classes, namely: structural analysis methods and added-content methods. The first class of methods mainly studies the structure of the social network, where the users present the nodes of the network. The links between the nodes vary according to the purpose of the analysis; they can be friendships, professional relationships or kinship ties. The second class groups the methods that are more interested in the content added by users than in the structure of the network.

Clearly, further research will be needed to cover in detail all social network analysis axis and methods. Several tasks in this context can be cited such as opinion and sentiment analysis or grouping like-minded users. On the basis of the findings presented in this paper, many issues in social network analysis need to be addressed and several methods and techniques could be improved.

References

1. Aci, M., İnan, C., Avci, M.: A hybrid classification method of k nearest neighbor, bayesian methods and genetic algorithm. *Expert Syst. Appl.* **37**(7), 5061–5067 (2010)
2. Aggarwal, C.C.: *Social Network Data Analytics*. Springer Publishing Company, Incorporated, 1st edn. (2011)
3. Aggarwal, C.C., Han, J., Wang, J., Yu, P.S.: A framework for clustering evolving data streams. In: *Proceedings of the 29th International Conference on Very Large Data Bases. VLDB '03*, vol. 29, pp. 81–92. VLDB Endowment, Germany (2003)

4. Ames, M., Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 971–980. ACM, USA (2007)
5. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Effects of user similarity in social media. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. pp. 703–712. WSDM '12, ACM, USA (2012)
6. Aouay, S., Jamoussi, S., Gargouri, F.: Feature based link prediction. In: 11th IEEE/ACS International Conference on Computer Systems and Applications, AICCSA. pp. 523–527. Qatar (2014)
7. Bader, B., Harshman, R., Kolda, T.: Temporal analysis of semantic graphs using asalsan. In: Seventh IEEE International Conference on Data Mining (ICDM). pp. 33–42. USA (2007)
8. Banerjee, A., Dhillon, I., Ghosh, J., Sra, S.: Expectation maximization for clustering on hyperspheres. Technical report, Univ. of Texas at Austin, USA (2003)
9. Beach, A., Gartrell, M., Xing, X., Han, R., Lv, Q., Mishra, S., Seada, K.: Fusing mobile, sensor, and social data to fully enable context-aware computing. In: Proceedings of the Eleventh Workshop on Mobile Computing Systems and Applications. pp. 60–65. HotMobile '10, ACM, USA (2010)
10. Cai, L., Hofmann, T.: Hierarchical document categorization with support vector machines. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management. pp. 78–87. CIKM '04, ACM, USA (2004)
11. Chang, M., Poon, C.K.: Using phrases as features in email classification. *J. Syst. Softw.* **82**(6), 1036–1045 (2009)
12. Chen, W., Wang, M.: A fuzzy c-means clustering-based fragile watermarking scheme for image authentication. *Expert Syst. Appl.* **36**(2), 1300–1307 (2009)
13. Chua, T., Tang, J., Hong, R., Li, H., Luo, Z., Zheng, Y.: Nus-wide: A real-world web image database from national university of singapore. In: Proceedings of the ACM International Conference on Image and Video Retrieval. pp. 48:1–48:9. CIVR '09, ACM, USA (2009)
14. Clauset, A., Moore, C., Newman, M.E.J.: Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98–101 (2008)
15. Conover, M., Goncalves, B., Ratkiewicz, J., Flammini, A., Menczer, F.: Predicting the political alignment of twitter users. pp. 192–199 (10 2011)
16. Cooper, M., Foote, J., Girgensohn, A., Wilcox, L.: Temporal event clustering for digital photo collections. *ACM Trans. Multimedia Comput. Commun. Appl.* **1**(3), 269–288 (2005)
17. Crandall, D.J., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: Proceedings of the 18th International Conference on World Wide Web. pp. 761–770. WWW '09, ACM, USA (2009)
18. Dai, W., Yang, Q., Xue, G., Yu, Y.: Self-taught clustering. In: Proceedings of the 25th International Conference on Machine Learning. pp. 200–207. ICML '08, ACM, USA (2008)
19. Dakiche, N., Tayeb, F.B.S., Slimani, Y., Benatchba, K.: Tracking community evolution in social networks: A survey. *Information Processing & Management* **56**(3), 1084–1102 (2019)
20. Daumé, III, J., Marcu, D.: Domain adaptation for statistical classifiers. *J. Artif. Int. Res.* **26**(1), 101–126 (2006)
21. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-fei, L.: Imagenet: A large-scale hierarchical image database. In: Computer Society Conference on Computer Vision and Pattern Recognition CVPR. pp. 248–255. USA (2009)

22. Dong, J., Zhao, Y., Peng, T.: Ontology classification for semantic-web-based software engineering. *IEEE Transactions on Services Computing* **2**, 303–317 (2009)
23. Du, J., Xian, Y., Yang, J.: A survey on social network visualization. In: Press, A. (ed.) *International Symposium on Social Science (ISSS 2015)*. pp. 275–279. China (2015)
24. Fortunato, S.: Community detection in graphs. *Physics reports* **486**(3-5), 75–174 (2010)
25. Fouss, F., Pirotte, A., Renders, J., Saerens, M.: Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Trans. on Knowl. and Data Eng.* **19**(3), 355–369 (2007)
26. Freeman, L.C.: Visualizing social networks. *Journal of Social Structure* **1** (2000)
27. Gallagher, A., Joshi, D., Yu, J., Luo, J.: Geo-location inference from image content and user tags. In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. pp. 55–62. USA (2009)
28. Girvan, M., Newman, M.E.: Community structure in social and biological networks. *Proceedings of the national academy of sciences* **99**(12), 7821–7826 (2002)
29. Goh, K.I., Oh, E., Kahng, B., Kim, D.: Betweenness centrality correlation in social networks. *Phys. Rev. E* **67**, 017101 (2003)
30. Hannachi, L., Asfari, O., Benblidia, N., Bentayeb, F., Kabachi, N., Boussaid, O.: Community extraction based on topic-driven-model for clustering users tweets. In: Zhou, S., Zhang, S., Karypis, G. (eds.) *Advanced Data Mining and Applications, 8th International Conference. Lecture Notes in Computer Science*, vol. 7713, pp. 39–51. Springer, China (2012)
31. Hays, J., Efros, A.A.: Im2gps: estimating geographic information from a single image. In: *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. pp. 1–8. IEEE Computer Society, USA (2008)
32. He, H., Wang, H., Yang, J., Yu, P.S.: Blinks: Ranked keyword searches on graphs. In: *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*. pp. 305–316. SIGMOD '07, ACM, USA (2007)
33. Ho, K.T., Bui, Q.V., Bui, M.: Dynamic social network analysis using author-topic model. In: *Innovations for Community Services - 18th International Conference, I4CS 2018, Slovakia, June 18-20, 2018, Proceedings*. pp. 47–62 (2018)
34. Irfan, R., King, C.K., Grages, D., Ewen, S., Khan, S.U., Madani, S.A., Kolodziej, J., Wang, L., Chen, D., Rayes, A., Tziritas, N., Xu, C., Zomaya, A.Y., Alzahrani, A.S., Li, H.: A survey on text mining in social networks. *The Knowledge Engineering Review* **30**, 157–170 (2015)
35. Jaffali, S., Jamoussi, S.: Principal component analysis neural network for textual document categorization and dimension reduction. In: *6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*,. pp. 835–839. IEEE, Tunisia (2012)
36. Jain, A.K.: Data clustering: 50 years beyond k-means. *Pattern Recogn. Lett.* **31**(8), 651–666 (2010)
37. Jo, T.: Neural text categorizer for exclusive text categorization. *International Journal of Information Science* **34**(1) (2010)
38. Joshi, D., Luo, J.: Inferring generic activities and events from image content and bags of geo-tags. In: *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval*. pp. 37–46. CIVR '08, ACM, USA (2008)
39. Kacholia, V., Pandit, S., Chakrabarti, S., Sudarshan, S., Desai, R., Karambelkar, H.: Bidirectional expansion for keyword search on graph databases. In: *Proceedings of the 31st International Conference on Very Large Data Bases*. pp. 505–516. VLDB '05, VLDB Endowment, Norway (2005)

40. Kamvar, S., Haveliwala, T., Manning, C., Golub, G.: Exploiting the block structure of the web for computing pagerank. Technical Report 2003-17, Stanford InfoLab, UK (2003)
41. Kashfia, S., Alhajj, R.: Emotion and sentiment analysis from twitter text. *Journal of Computational Science* (2019)
42. Kavitha, V., Punithavalli, M.: Clustering time series data stream - A literature survey. *international journal for computer science and information security ijsis* **8**(1), 289–294 (2010)
43. Khalessizadeh, S.M., Zaefarian, R., Nasser, S., Ardil, E.: Genetic mining: Using genetic algorithm for topic based on concept distribution. *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering* **2**(1), 35 – 38 (2008)
44. Khemakhem, I.T., Jamoussi, S., Hamadou, A.B.: POS tagging without a tagger: Using aligned corpora for transferring knowledge to under-resourced languages. *Computación y Sistemas* **20**(4) (2016)
45. Leicht, E.A., Holme, P., Newman, M.E.J.: Vertex similarity in networks. *Phys. Rev. E* **73**(2), 026–120 (2006)
46. Li, X., Chen, H.: Recommendation as link prediction in bipartite graphs. *Decis. Support Syst.* **54**(2), 880–890 (2013)
47. Li, Z.L., Fang, X., Sheng, O.R.L.: A survey of link recommendation for social networks: Methods, theoretical foundations, and future research directions. *ACM Trans. Manage. Inf. Syst.* **9**(1), 1:1–1:26 (Oct 2017)
48. Lin, Y., Chi, Y., Zhu, S., Sundaram, H., Tseng, B.L.: Facetnet: A framework for analyzing communities and their evolutions in dynamic networks. In: *Proceedings of the 17th International Conference on World Wide Web*. pp. 685–694. WWW '08, ACM, USA (2008)
49. Liu, D., Hua, X.S., Yang, L., Wang, M., Zhang, H.: Tag ranking. In: *Proceedings of the 18th International Conference on World Wide Web*. pp. 351–360. WWW '09, ACM, USA (2009)
50. Liu, H., Hu, Z., Haddadi, H., Tian, H.: Hidden link prediction based on node centrality and weak ties. *EPL (Europhysics Letters)* **101**(1) (2013)
51. Lü, L., Jin, C., Zhou, T.: Similarity index based on local paths for link prediction of complex networks. *Phys. Rev. E* **80**, 046–122 (2009)
52. Macskassy, S.A., Provost, F.: A simple relational classifier. In: *Proceedings of the Second Workshop on Multi-Relational Data Mining (MRDM-2003) at KDD-2003*. pp. 64–76 (2003)
53. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. *Annual review of sociology* **27**(1), 415–444 (2001)
54. Menon, A.K., Elkan, C.: Link prediction via matrix factorization. In: *Proceedings of the 2011 European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part II*. pp. 437–452. ECML PKDD'11, Springer-Verlag, Germany (2011)
55. Miao, D., Duan, Q., Zhang, H., Jiao, N.: Rough set based hybrid algorithm for text classification. *Expert Syst. Appl.* **36**(5), 9168–9174 (2009)
56. Moody, J., McFarland, D., Bender-deMoll, S.: Dynamic network visualization. *American Journal of Sociology* **110**(4), 1206–1241 (2005)
57. Moradabadi, B., Meybodi, M.R.: Link prediction in weighted social networks using learning automata. *Engineering Applications of Artificial Intelligence* **70**, 16–24 (2018)

58. Naaman, M., Harada, S., Wang, Q., Garcia-Molina, H., Paepcke, A.: Context data in geo-referenced digital photo collections. In: *Proceedings of the 12th Annual ACM International Conference on Multimedia*. pp. 196–203. MULTIMEDIA '04, ACM, USA (2004)
59. Neville, J., Jensen, D.: Leveraging relational autocorrelation with latent group models. In: *Fifth IEEE International Conference on Data Mining (ICDM'05)*. pp. 8–pp. IEEE (2005)
60. Newman, M.E.J.: Clustering and preferential attachment in growing networks. *Phys. Rev. E* **64**, 025–102 (2001)
61. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. on Knowl. and Data Eng.* **22**(10) (2010)
62. Papadimitriou, A., Symeonidis, P., Manolopoulos, Y.: Fast and accurate link prediction in social networking systems. *J. Syst. Softw.* **85**(9), 2119–2132 (2012)
63. Qi, G., Aggarwal, C.C., Huang, T.S.: Community detection with edge content in social media networks. In: Kementsietsidis, A., Salles, M.A.V. (eds.) *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*. pp. 534–545. IEEE Computer Society, USA (2012)
64. Qi, G., Hua, X., Zhang, H.: Learning semantic distance from community-tagged media collection. In: *Proceedings of the 17th ACM International Conference on Multimedia*. pp. 243–252. MM '09, ACM, USA (2009)
65. Raina, R., Battle, A., Lee, H., Packer, B., Ng, A.Y.: Self-taught learning: Transfer learning from unlabeled data. In: *Proceedings of the 24th International Conference on Machine Learning*. pp. 759–766. ICML '07, ACM, USA (2007)
66. Rao, Y., Li, X.: A topic-based dynamic clustering algorithm for text stream. In: *International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2015)*. pp. 480–483. Thailand (2015)
67. Rocchio, J.J.: Relevance feedback in information retrieval. pp. 313–323 (1971)
68. Rossetti, G., Cazabet, R.: Community discovery in dynamic networks: A survey. *ACM Comput. Surv.* **51**(2), 35:1–35:37 (Feb 2018)
69. Sapountzi, A., Psannis, K.E.: Social networking data analysis tools and challenges. *Future Generation Computer Systems* **86**, 893 – 913 (2018)
70. Seifzadeh, S., Farahat, A.K., Kamel, M.S., Karray, F.: Short-text clustering using statistical semantics. In: Gangemi, A., Leonardi, S., Panconesi, A. (eds.) *Proceedings of the 24th International Conference on World Wide Web*. pp. 805–810. ACM, USA (2015)
71. Shen, Z., Ma, K.: Mobivis: A visualization system for exploring mobile data. In: *Proceedings of MobiVis: A Visualization System for Exploring Mobile Data*. pp. 175–182. IEEE, Japan (2008)
72. Shen, Z., Ma, K., Eliassi-Rad, T.: Visual analysis of large heterogeneous social networks by semantic and structural abstraction. *IEEE Transactions on Visualization and Computer Graphics* **12**(6), 1427–1439 (2006)
73. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: *Proceedings of the 17th International Conference on World Wide Web*. pp. 327–336. WWW '08, ACM, USA (2008)
74. Sun, J., Tao, D., Faloutsos, C.: Beyond streams and graphs: Dynamic tensor analysis. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 374–383. KDD '06, ACM, USA (2006)
75. Tang, F., Gao, Y.: Fast near duplicate detection for personal image collections. In: *Proceedings of the 17th ACM International Conference on Multimedia*. pp. 701–704. MM '09, ACM, USA (2009)

76. Tang, J., Yan, S., Hong, R., Qi, G., Chua, T.: Inferring semantic concepts from community-contributed images and noisy tags. In: Proceedings of the 17th ACM International Conference on Multimedia. pp. 223–232. MM '09, ACM, USA (2009)
77. Valverde-Rebaza, J.C., de Andrade Lopes, A.: Exploiting behaviors of communities of twitter users for link prediction. *Social Netw. Analys. Mining* **3**(4), 1063–1074 (2013)
78. Wang, X.J., Zhang, L., Li, X., Ma, W.Y.: Annotating images by mining image search results. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(11), 1919–1932 (2008)
79. Wang, Z., Song, Y., Zhang, C.: Transferred Dimensionality Reduction, pp. 550–565. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
80. Weinberger, K.Q., Slaney, M., Van Zwol, R.: Resolving tag ambiguity. In: Proceedings of the 16th ACM International Conference on Multimedia. pp. 111–120. MM '08, ACM, USA (2008)
81. Wu, P., Tretter, D.: Close & closer: social cluster and closeness from photo collections. In: Gao, W., Rui, Y., Hanjalic, A., Xu, C., Steinbach, E.G., El-Saddik, A., Zhou, M.X. (eds.) *ACM Multimedia*. pp. 709–712. ACM, USA (2009)
82. Wu, S., Sun, J., Tang, J.: Patent partner recommendation in enterprise social networks. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. pp. 43–52. WSDM '13, ACM, USA (2013)
83. Xiang, R., Neville, J.: Collective inference for network data with copula latent markov networks. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. pp. 647–656. WSDM '13, ACM, New York, NY, USA (2013)
84. Xu, G., Zhang, Y., Li, L.: *Web Mining and Social Networking: Techniques and Applications*. Springer-Verlag New York, Inc., 1st edn. (2011)
85. Xu, X., Zhang, F., Niu, Z.: An ontology-based query system for digital libraries. In: Pacific-Asia Workshop on Computational Intelligence and Industrial Application. pp. 222–226. China (2008)
86. Yamamoto, T., Honda, K., Notsu, A., Ichihashi, H.: A comparative study on tibia imputation methods in fcmdd-based linear clustering with relational data. *Adv. Fuzzy Sys.* **2011**, 265170:1–265170:10 (2011)
87. Yu, J., Luo, J.: Leveraging probabilistic season and location context models for scene understanding. In: Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval. pp. 169–178. CIVR '08, ACM, USA (2008)
88. Zadrozny, B.: Learning and evaluating classifiers under sample selection bias. In: Proceedings of the Twenty-first International Conference on Machine Learning. pp. 114–121. ICML '04, ACM, USA (2004)