Computer Science Review 37 (2020) 100286

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Survey

Community detection in node-attributed social networks: A survey

Petr Chunaev

National Center for Cognitive Technologies, ITMO University, Saint Petersburg, Russia

ARTICLE INFO

Article history:
Received 20 December 2019
Received in revised form 29 May 2020
Accepted 11 July 2020
Available online xxxx

Keywords:
community detection
social network
complex network
node-attributed graph
clusterization

ABSTRACT

Community detection is a fundamental problem in social network analysis consisting, roughly speaking, in unsupervised dividing social actors (modeled as nodes in a social graph) with certain social connections (modeled as edges in the social graph) into densely knitted and highly related groups with each group well separated from the others. Classical approaches for community detection usually deal only with the structure of the network and ignore features of the nodes (traditionally called node attributes), although the majority of real-world social networks provide additional actors' information such as age, gender, interests, etc. It is believed that the attributes may clarify and enrich the knowledge about the actors and give sense to the detected communities. This belief has motivated the progress in developing community detection methods that use both the structure and the attributes of the network (modeled already via a node-attributed graph) to yield more informative and qualitative community detection results.

During the last decade many such methods based on different ideas and techniques have appeared. Although there exist partial overviews of them, a recent survey is a necessity as the growing number of the methods may cause repetitions in methodology and uncertainty in practice.

In this paper we aim at describing and clarifying the overall situation in the field of community detection in node-attributed social networks. Namely, we perform an exhaustive search of known methods and propose a classification of them based on when and how the structure and the attributes are fused. We not only give a description of each class but also provide general technical ideas behind each method in the class. Furthermore, we pay attention to available information which methods outperform others and which datasets and quality measures are used for their performance evaluation. Basing on the information collected, we make conclusions on the current state of the field and disclose several problems that seem important to be resolved in future.

© 2020 Elsevier Inc. All rights reserved.

Contents

E-mail address: chunaev@itmo.ru.

https://doi.org/10.1016/j.cosrev.2020.100286
1574-0137/© 2020 Elsevier Inc. All rights reserved.

1. Introduction

Community detection is a fundamental problem in social network analysis consisting, roughly speaking, in unsupervised dividing social actors into densely knitted and highly related groups with each group well separated from the others. One class of classical community detection methods mainly deal only with the *structure* of social networks (i.e. with connections between social actors) and ignore actors' features. There exist a variety of such structure-aware methods that have shown their efficiency in multiple applications (see [65,112,117]). However, the majority of real-world social networks provide more information about social actors than just connections between them. Indeed, it is rather common that certain actors' *attributes* such as age, gender, interests, etc., are available. When it is so, the social network is called *node-attributed* (recall that the actors are represented via nodes). According to [191], attributes form the second dimension, besides the structural one, in social network representation. There is another class of classical community detection methods (being opposite to the structure-aware ones, in a sense) that use only node attributes to detect communities and completely ignore connections between social actors. A representative of the attributes-aware methods is well-known k-means clustering algorithm taking attribute vectors as an input. Clearly, methods that deal only with structure or only with attributes do not use all the information available in a node-attributed social network. Naturally, this issue can be overcome if a method would somehow jointly use structure and attributes while detecting communities. Developing of such methods became a novel field in social network analysis [24]. The field is moreover promising as the joint usage is believed to clarify and enrich the knowledge about social actors and to describe the powers that form their communities [24].

During the last decade numerous methods based on different ideas and techniques have appeared in the field. Although there exist some partial overviews of them, especially in Related Works sections of published papers and in the survey [24] published in 2015, a recent summary of the subject is a necessity as the growing number of the methods may cause repetitions in methodology and uncertainty in practice.

In this survey, we aim at describing and clarifying the overall situation in the field. Namely, we perform an exhaustive search of existing community detection methods for node-attributed social networks. What is more, we propose a classification of them based on when and how they use and fuse network structure and attributes. We not only give a description of each class but also provide general technical ideas behind each method in the class. Furthermore, we pay attention to available information which methods outperform others and which datasets and quality measures are used for their performance evaluation. Basing on the information collected, we make conclusions on the current state of the field and disclose several problems that seem important to be resolved in future.

To be more precise, let us describe the content of the survey. In Section 2, we first provide the reader with the notation used in the survey and state the problem of community detection in node-attributed social networks. We further briefly discuss the traditional argumentation in support of such a community detection and the effect of fusing network structure and attributes. In Section 3, we give information about the related survey works and explain how the search of relevant literature was organized in our case. We also indicate which methods are included in the survey and which are not. Additionally, we explain why references throughout the survey are made in a certain way. Section 4 introduces the classification that we propose for community detection methods under consideration. In Section 5, we discuss the most popular datasets and quality measures for evaluation of community detection results. This section is also helpful for simplifying exposition in the forthcoming sections. Sections 6–8 contain descriptions of the classes of methods and their representatives. In Section 9, we analyze the overall situation in the field basing on the information from Sections 6–8. Among other things, we disclose several methodological problems that are important to resolve in future studies, in our opinion. Our conclusions on the topic are summarized in Section 10.

2. Community detection problem for node-attributed social networks and the effect of fusing network structure and attributes

2.1. Necessary notation and the community detection problem statement

We represent a node-attributed social network as triple (*node-attributed graph*) $G = (\mathscr{V}, \mathscr{E}, \mathscr{A})$, where $\mathscr{V} = \{v_i\}$ is the set of nodes (vertices) representing social actors, $\mathscr{E} = \{e_{ij}\}$ the set of edges representing connections between the actors (e_{ij} stands for the edge between nodes v_i and v_j), and $\mathscr{A}$ the set of attribute vectors $A(v_i) = \{a_k(v_i)\}$ associated with nodes in $\mathscr{V}$ and containing information about actors' features. Furthermore, $|\mathscr{V}| = n$, $|\mathscr{E}| = m$ and the dimension of the attribute vectors is d. The domain of a_k,

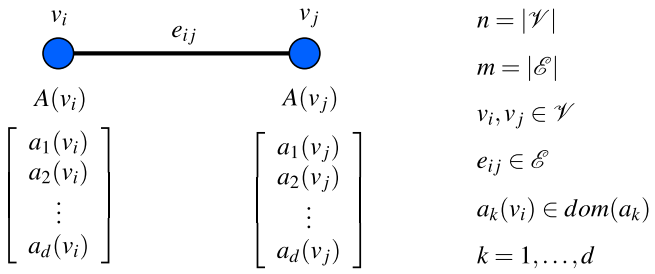

Fig. 1. Notation related to the triple (node-attributed graph) $G = (\mathscr{V}, \mathscr{E}, \mathscr{A})$, where $\mathscr{V} = \{v_i\}$, $\mathscr{E} = \{e_{ij}\}$ and $\mathscr{A} = \{A(v_i)\}$.

i.e. the set of possible values of kth element of attribute vectors $a_k(v_i)$, is denoted by $dom(a_k)$. In these terms, kth attribute of node v_i is referred to as $a_k(v_i)$. The notation introduced above is summarized in Fig. 1. Note that pairs $(\mathscr{V}, \mathscr{E})$ and $(\mathscr{V}, \mathscr{A})$ are correspondingly called the *structure* (or *topology*) and the *attributes* (or *semantics*) of node-attributed graph G.

By *community detection*[1] *in node-attributed graph* $G = (\mathscr{V}, \mathscr{E}, \mathscr{A})$ we mean *unsupervised* partitioning the set of nodes $\mathscr{V}$ into N subsets (*communities* or *clusters*) $C_k \subset \mathscr{V}$, with $C = \{C_k\}_{k=1}^N$ called a *partition*, such that $\mathscr{V} = \bigcup_{k=1}^N C_k$ and a certain balance between the following two properties is achieved:

(a) *structural closeness*, i.e. nodes within a community are structurally close to each other, while nodes in different communities are not;
(b) *attribute homogeneity*, i.e. nodes within a community have homogeneous attributes, while nodes in different communities do not.

Since one can meet variations of the above-mentioned definitions in the relevant literature, it is worth giving several comments on them. First, the number of communities N can be either known in advance or determined during the community detection process automatically. Second, the communities C_k may be defined to be disjoint or overlapping. Third, the property $\mathscr{V} = \bigcup_{k=1}^N C_k$ is sometimes omitted if the resulting partition is not required to include all the nodes from $\mathscr{V}$. Fourth, the notion of structural closeness and attribute homogeneity may seem vague at the moment but hopefully become more evident after Sections 2.2 and 4 where reasons and particular measures for them are discussed. Fifth, the definitions given are for the case of nodes and edges each of one type. It is of course possible that social actors and connections between them can be of different types in a social network and thus one should take this heterogeneity into account. This situation is however closer to the notion of multilayer networks that are out of scope of the present survey (see also Section 3).

2.2. Structural closeness and attribute homogeneity. The effect of fusing structure and attributes

The structural closeness requirement is based on the recent concepts of a (structural) community in a social network. For example, communities are thought in [72] as subsets of nodes with dense connections within the subsets and sparse in between. In its turn, [145] adopts the intuition that nodes within the same community should be better connected than they would be by chance. A possible measure for that is famous Newman's Modularity [143] that has become an influential tool for structure-based community detection in social networks [24,33]. Multiple Modularity modifications and other measures have been also proposed to assess structural closeness [33]. In fact, the precise meaning of structural closeness in each community detection method is determined by the measure chosen.

The attribute homogeneity requirement is based on the social science founding (see e.g. [64,110,129,131]) that actors' features can reflect and affect community structure in social networks. The well-known principle of homophily in social networks states that like-minded social actors have a higher likelihood to be connected [131]. Thus community detection process taking into account attribute homogeneity may provide results of better quality [24]. Oppositely to the situation with structural closeness measures, the attribute homogeneity is usually measured by Entropy that quantifies the degree of disorder of attribute vectors in $(\mathscr{V}, \mathscr{A})$ within the communities detected.

Let us now discuss different points of view on the effect of fusing structure and attributes. From one side, multiple experiments, e.g. in [41,71,135,168,205] and many other papers cited in this survey, suggest that the structure and the attributes of a node-attributed social network often provide complementary information that improves community detection quality. For example, attributes may compensate the structural sparseness of a real-world social network [102,204], while structural information may be helpful in resolving the problem of missing or noisy attributes [102,168,204]. What is more, it is observed in [56] that structure-only or attributes-only community detection is often not as effective as when both sources of information are used. From the other side, some experiments (see e.g. [5,211]) suggest that this is not always true, and network structure and attributes may be orthogonal and contradictory thus leading to ambiguous community detection results. Moreover, relations between these sources of information may be highly non-linear and challenging to analyze [187,204].

Besides the above-mentioned points, our general impression is that there is no widely accepted opinion on the effect of fusing structure and attributes and how this fusion can influence community detection quality. Let us illustrate this with an example. Suppose that the structure of a certain node-attributed social network is ideally correlated with a half of attributes and is wholly orthogonal to another half. For simplicity, let the dimension of attribute vectors be small so that there is no sense to fight against the curse of dimensionality. Now we follow the popular suggestion that the mismatch between structure and attributes negatively affects community detection quality [132] and that the existence of structure–attributes correlation offers "a unique opportunity to improve the learning performance of various graph mining tasks" [119]. The choice is clear then: we need to use the structure and only the ideally correlated attributes for community detection. It turns out however that we are going to use two sources of information that mostly duplicate each other. Why should we expect that this improves the quality of detected communities? From our side, we would presume that the structure and the chosen half of attributes (considered separately or jointly) would yield very similar or even the same communities, with all the ensuing consequences for assessing their quality. Should not we use just one of the sources then? Furthermore, it is not clear to us why the other half of attributes should be omitted. Generally speaking, they may contain valuable information for community detection and thus omitting them because of the lack of correlation with the structure is rather questionable.

In any case, a focused theoretical study of when the fusion of structure and attributes is worthy and when not for community detection (ideally, in terms of subclasses of node-attributed social networks) seems to be an extremely important thing that would allow to remove the above-mentioned contradictions.

[1] It is also called "community discovery" or "clusterization".

3. Related works and processing the relevant literature

3.1. Related works

There is a variety of surveys and comparative studies considering community detection in social networks without attributes, in particular, [44,65,166,200]. At the same time, the survey [24] seems to be the only one on community detection in node-attributed social networks. Obviously, since it was published in 2015, many new methods adapting different fusion techniques have appeared in the field. Furthermore, a big amount of the methods that had been proposed before 2015 are not covered by [24], in particular, some based on objective function modification, non-negative matrix factorization, probabilistic models, clustering ensembles, etc. In a sense, the technique-based classification of methods in [24] is also sometimes confusing. For example, CODICIL [160], a method based on assigning attribute-aware weights on graph edges, is not included in [24, Section 3.2. Weight modification according to node attributes], but to [24, Section 3.7. Other methods]. Although [24] is a well-written and highly cited paper, a recent survey of community detection methods for node-attributed social networks is clearly required.

Besides [24], almost every paper on the topic contains a Related Works section. It typically has a short overview of preceding methods and an attempt to classify them. We observe that many authors are just partly aware of the relevant literature and this sometimes leads to repetitions in approaches. Furthermore, multiple classifications (usually technique-based) are mainly not full and even contradictory.

3.2. Relevant literature search process

At the beginning, we started the search of relevant literature using regular and scientific search engines making the queries like "community detection" or "clustering" or "community discovery" in "node-attributed social networks" or "node-attributed graphs". Within the search process it became evident that other queries also lead to the relevant literature. In particular, "clustering an attribute-aware graph", "community detection in networks with node attributes", "description-oriented community detection", "semantic clustering of social networks", "structure and attributes community detection", "joint cluster analysis of attribute and relationship data", "community discovery integrating links and tags", "attributed graph partitioning", "node attribute-enhanced community detection", "community detection based on structure and content", etc. It can be also observed that node-attributed networks and graphs are also sometimes called "augmented networks", "graphs with feature vectors", "feature-rich networks" and "multi-attributed graphs". This variety of terms suggests that there is still no established terminology in the field and emphasizes the significance of our survey, where we try to use consistent terminology.

After the above-mentioned exhaustive search, we learned the references in the found papers. Among other things, it brought us to ideologically close papers devoted, for example, to "attributed information networks", "annotated document networks", "multi-layer networks" and "subspace-based clustering". We stopped further search when we could not find any new relevant references. Since this happened in the middle of 2019, the survey covers the found papers that had been published in journals or conference proceedings before this date.

3.3. The format of references to methods and datasets

It turns out that several methods for community detection in node-attributed social networks can be proposed in one paper. Therefore, a regular reference of the form [ReferenceNumber] may be not informative enough. From the other hand, authors usually provide their methods with short names like SA-Cluster, CODICIL or CESNA.[2] Some of the names are rather familiar to researches in the field. Thus it seems reasonable to make a reference of the form MethodName [ReferenceNumber] and so we do in what follows. However, not all the methods that are mentioned in the survey are included in our classification and thus discussed in a more detailed manner (as some are just out of scope). To distinguish the cases, we write names of the methods included in our classification in **bold**, e.g. **SCMAG** [98], **UNCut** [205] and **DCM** [158]. Such a format means that the reader can find short descriptions of the methods **SCMAG** [98], **UNCut** [205] and **DCM** [158] in our survey. References like DB-CSC [81], FocusCO [155] and ACM [193] mean that the corresponding methods are not included in our classification. In this case the reader is recommended to go directly to the papers [81, 155] and [193] to get additional information.

A similar scheme is applied to names of the node-attributed network datasets discussed in Section 5 and used in further classification. The reader is assumed to have in mind that if a dataset name is written in **bold**, then its description can be found in Tables 1, 2 or 3. Note also that various versions of the datasets from Tables 1, 2 or 3 are in fact used in different papers. To show that a dataset is somehow different from the description in Tables 1, 2 or 3, we mark it by *. For example, a DBLP dataset with the number of nodes and edges different from **DBLP10K** and **DBLP84K** in Table 2 is denoted by **DBLP*** in Sections 6–8.

3.4. Note on multi-layer network clustering

In the survey, we do not consider community detection methods for multi-layer networks, where different types of vertices and edges may present at different layers [99,109]. Nevertheless, we mention some of these methods from time to time in corresponding remarks. It is though important to note that node-attributed networks of different nature may be clearly considered as a particular case of multi-layer ones. In the majority of papers covered by the present survey, this connection is however rarely commented.

Let us also emphasize that multi-layer networks (graphs) require special analysis taking into account the heterogeneity of vertices and edges on different layers. A separate survey and an extensive comparable study of such methods is an independent and useful task (see partial overviews e.g. in [25,99,109]).

3.5. Note on subspace-based clustering

Following the above-mentioned definition of community detection in node-attributed social networks, we mainly consider in the survey the methods that can use the full attribute space and find communities covering the whole network. However, there is a big class of special methods that explore subspaces of attributes and/or deal with subgraphs of an initial graph, e.g. GAMer [82,83], DB-CSC [81], SSCG [84], FocusCO [155] and ACM [193]. The main idea behind the subspace-based (also known as projection-based) clustering methods is that not all available semantic information (attributes) is relevant to obtain good-quality communities [79, 80]. For this reason, one has to somehow choose the appropriate

[2] If this is not the case, we allowed ourselves to invent our own names suggesting the class the method belongs to in our classification.

Table 1
Most popular small size datasets.

Dataset	Description	Source
Political Books	All books in this dataset were about U.S. politics published during the 2004 presidential election and sold by Amazon.com. Edges between books means two books are always bought together by customers. Each book has only one attribute termed as political persuasion, with three values: (1) conservative; (2) liberal; and (3) neutrality	Link
WebKB	A classified network of 877 webpages (nodes) and 1608 hyperlinks (edges) gathered from four different universities Web sites (Cornell, Texas, Washington, and Wisconsin). Each web page is associated with a binary vector, whose elements take the value 1 if the corresponding word from the vocabulary is present in that webpage, and 0 otherwise. The vocabulary consists of 1703 unique words. Nodes are classified into five classes: course, faculty, student, project, or staff.	Link [45]
Twitter	A collection of several tweet networks: (1) Politics-UK dataset is collected from Twitter accounts of 419 Members of Parliament in the United Kingdom in 2012. Each user has 3614-dimensional attributes, including a list of words repeated more than 500 times in their tweets. The accounts are assigned to five disjoint communities according to their political affiliation. (2) Politics-IE dataset is collected from 348 Irish politicians and political organizations, each user has 1047-dimensional attributes. The users are distributed into seven communities. (3) Football dataset contains 248 English Premier League football players active on Twitter which are assigned to 20 disjoint communities, each corresponding to a Premier League club. (4) Olympics dataset contains users of 464 athletes and organizations involved in the London 2012 Summer Olympics. The users are grouped into 28 disjoint communities, corresponding to different Olympic sports.	Link 1 Link 2 [73]
Lazega	A corporate law partnership in a Northeastern US corporate law firm; possible attributes: (1: partner; 2: associate), office (1: Boston; 2: Hartford; 3: Providence); 71 nodes and 575 edges	[114]
Research	A research team of employees in a manufacturing company; possible attributes: location (1: Paris; 2: Frankfurt; 3: Warsaw; 4: Geneva), tenure (1: 1–12 months; 2: 13–36 months; 3: 37–60 months; 4: 61+ months); 77 nodes and 2228 edges	[46]
Consult	Network showing relationships between employees in a consulting company; possible attributes: organizational level (1: Research Assistant; 2: Junior Consultant; 3: Senior Consultant; 4: Managing Consultant; 5: Partner), gender (1: male; 2: female); 46 nodes and 879 edges	[46]

attribute subspace to avoid the so-called *curse of dimensionality* (see [24, Section 3.2]) and reveal significant communities that would not be detected if all available attributes were considered.

To be precise, some of the methods that we discuss below partly use this idea, e.g. **WCru** [48,49] (cf. the definition of a *point of view* in the papers), **DVil** [186], **SCMAG** [98], **UNCut** [205], **DCM** [158], etc., but still can work with the full attribute space. In any case, a separate survey on subspace-based methods for community detection in node-attributed social networks would be a very valuable complement to the current one.

3.6. Note on community detection in node-attributed networks of different type and its applications

Clearly, community detection tools for node-attributed social networks are suitable for networks of different nature. That is why, besides obvious applications in marketing (recommender systems, targeted advertisements and user profiling) [8], the tools are used for search engine optimization and spam detection [138,160], in counter-terrorist activities and disclosing fraudulent schemes [138]. They are also applied to analysis of protein–protein interactions, genes and epidemics [138].

Another possible application is document network clustering. Note that such a clustering is historically preceding to community detection in node-attributed social networks and is rich methodologically on its own [3,139,162]. One should take into account however that social communities although have similar formal description with document clusters have inner and more complicated forces to be formed and act. What is more, it has been shown that methods for community detection in node-attributed social networks outperform preceding methods for document network clustering in many cases, see [14,202,204, 212]. For this reason, we do not consider document network clustering methods in this survey.

4. Our classification of community detection methods for node-attributed social networks

In previous works, the classification of methods for community detection in node-attributed social networks was done mostly with respect to the techniques used (e.g. distance-based or random walk-based). We partly follow this principle but at a lower level. At the upper level we group the methods by when structure and attributes are fused with respect to the community detection process, see Fig. 2. Namely, we distinguish the classes of

- *early fusion methods* that fuse structure and attributes before the community detection process,
- *simultaneous fusion methods* that fuse structure and attributes simultaneously with the community detection process,
- *late fusion methods* that first partition structure and attributes separately and further fuse the partitions obtained.

Such a classification allows an interested researcher/data scientist to estimate the labor costs of software implementation of the method chosen for use in practice. Indeed, early fusion methods require just a preprocessing (fusion) procedure converting information about structure and attributes into a format that is often suitable for classical community detection algorithms. For example, weight-based early fusion methods convert attribute vectors into the graph form and further merge the structure- and attributes-aware graphs into a weighted graph that can be processed by graph clustering algorithms with existing implementations. Implementation of late fusion methods is also rather simple. Namely, at the first step they detects communities by classical graph and vector clustering algorithms applied to structure and attributes separately. At the second step, the communities obtained are somehow merged by an existing (or at least easy-implemented) algorithm. Oppositely to the early and late fusion methods, simultaneous fusion ones require more programmer work as usually assume either a completely new implementation or essential modifications to existing ones. The situation is aggravated by the fact that their source codes are rarely available.

As seen from Fig. 2, we also divide the methods within each class into technique-used subclasses. Let us emphasize that the priority in this division is the fusion technique. For example, by "weight-based methods" we mean those which form a weighted graph while fusing structure and attributes. The majority of such methods further use graph clustering algorithms for community detection (and this is reasonable). However, some may still transform the graph into the distance matrix form and then use distance-based clustering algorithms. Such methods are

Table 2
Most popular medium size datasets.

Dataset	Description	Source
Political Blogs	A non-classified network of 1,490 webblogs (nodes) on US politics with 19,090 hyperlinks (edges) between the webblogs. Each node has an attribute describing its political leaning as either liberal or conservative (represented by 0 and 1).	Link [2]
DBLP10K	A non-classified co-author network extracted from DBLP Bibliography (four research areas of database, data mining, information retrieval and artificial intelligence) with 10,000 authors (nodes) and their co-author relationships (edges). Each author is associated with two relevant categorical attributes: prolific and primary topic. For attribute "prolific", authors with $\geqslant 20$ papers are labeled as highly prolific; authors with > 10 and < 20 papers are labeled as prolific and authors with $\leqslant 10$ papers are labeled as low prolific. Node-attribute values for "primary topic" (100 research topics) are obtained via topic modeling. Each extracted topic consists of a probability distribution of keywords which are most representative of the topic.	Link [212]
DBLP84K	A larger non-classified co-author network extracted from DBLP Bibliography (15 research areas of database, data mining, information retrieval, artificial intelligence, machine learning, computer vision, networking, multimedia, computer systems, simulation, theory, architecture, natural language processing, human–computer interaction, and programming language) with 84,170 authors (nodes) and their co-author relationships (edges). Each author is associated with two relevant categorical attributes: prolific and primary topic, defined in a similar way as in DBLP10.	Link [212]
Cora	A classified network of machine learning papers with 2,708 papers (nodes) and 5,429 citations (edges). Each node is attributed with a 1433-dimension binary vector indicating the absence/presence of words from the dictionary of words collected from the corpus of papers. The papers are classified into 7 subcategories: case-based reasoning, genetic algorithms, neural networks, probabilistic methods, reinforcement learning, rule learning and theory.	Link 1 Link 2 [167]
CiteSeer	A classified citation network in the field of machine learning with 3,312 papers (nodes) and 4,732 citations (edges). Each node is attributed with a binary vector indicating the absence/presence of the corresponding words from the dictionary of the 3,703 words collected from the corpus of papers. Papers are classified into 6 classes.	Link 1 Link 2 [167]
Sinanet	A classified microblog user relationship network extracted from the sina-microblog website (http://www.weibo.com) with 3,490 users (nodes) and 30,282 relationships (edges). Each node is attributed with 10-dimensional numerical attributes describing the interests of the user.	Link [102]
PubMed Diabetes	A classified citation networks extracted from the PubMed database pertaining to diabetes. It contains 19,717 publications (nodes) and 44,338 citations (edges). Each node is attributed by a TF–IDF weighted word vector from a dictionary that consists of 500 unique words.	Link
Facebook100	A non-classified Facebook users network with 6,386 users (nodes) and 435,324 friendships (edges). The network is gathered from Facebook users of 100 colleges and universities (e.g. Caltech, Princeton, Georgetown and UNC Chapel Hill) in September 2005. Each user has the following attributes: ID, a student/faculty status flag, gender, major, second major/minor (if applicable), dormitory(house), year and high school.	Link [184,185]
ego-Facebook	Dataset consists of 'circles' ('friends lists') from Facebook with 4039 nodes and 88234 edges. Facebook data was collected from survey participants using a Facebook app. The dataset includes node features (profiles), circles, and ego networks.	Link [118]
LastFM	A network gathered from the online music system Last.fm with 1,892 users (nodes) and 12,717 friendships on Last.fm (edges). Each node has 11,946-dimensional attributes, including a list of most listened music artists, and tag assignments.	Link
Delicious	A network of 1,861 nodes, 7,664 edges and 1,350 attributes. This is a publicly available dataset from the HetRec 2011 workshop that has been obtained from the Delicious social bookmarking system. Its users are connected in a social network generated from Delicious mutual fan relations. Each user has bookmarks, tag assignments, that is, [user, tag, bookmark] tuples, and contact relations within the social network. The tag assignments were transformed to attribute data by taking all tags that a user ever assigned to any bookmark and assigning those to the user.	Link
Wiki	A network with nodes as web pages. The link among different nodes is the hyperlink in the web page. 2,405 nodes, 12,761 edges, 4,973 attributes, 17 labels	Link
ego-Twitter	This dataset consists of 'circles' (or 'lists') from Twitter. Twitter data was crawled from public sources. The dataset includes node features (profiles), circles, and ego networks. Nodes 81306, Edges 1768149	Link [118]

Table 3
Most popular large size datasets.

Dataset	Description	Source
Flickr	A network with 100,267 nodes, 3,781,947 edges and 16,215 attributes collected from the internal database of the popular Flickr photo sharing platform. The social network is defined by the contact relation of Flickr. Two vertices are connected with an undirected edge if at least one undirected edge exists between them. Each user has a list of tags associated that he/she used at least five times. Tags are limited to those used by at least 50 users. Users are limited to those having a vocabulary of more than 100 and less than 5,000 tags.	Link [160]
Patents	A patent citation network with vertices representing patents and edges depicting the citations between. A subgraph containing all the patents from the year 1988 to 1999. Each patent has six attributes, grant year, number of claims, technological category, technological subcategory, assignee type, and main patent class. There are 1,174,908 vertices and 4,967,216 edges in the network.	Link 1 Link 2
ego-G+	This dataset consists of 'circles' from Google+. Google+ data was collected from users who had manually shared their circles using the 'share circle' feature. The dataset includes node features (profiles), circles, and ego networks. Nodes 107,614, Edges 13,673,453. Each node has four features: job title, current place, university, and workplace. A user-pair(edge) is compared using knowledge graphs based on, Category: Occupations, Category:Companies by country and industry, Category: Countries, Category:Universities and colleges by country.	Link [118]

still called "weight-based". One more example: "distance-based methods" are called in this way as directly produce a distance matrix while fusing structure and attributes, independently on how this matrix is further processed.

What is more, we provide short descriptions of methods in each class and subclass in tables in Sections 6–8. In particular, we briefly describe the community detection algorithm and its input used in the method, and the type of communities obtained (overlapping or not). Furthermore, we mention which datasets and quality measured are used by method's authors for evaluation of community detection results. In addition, for each method we

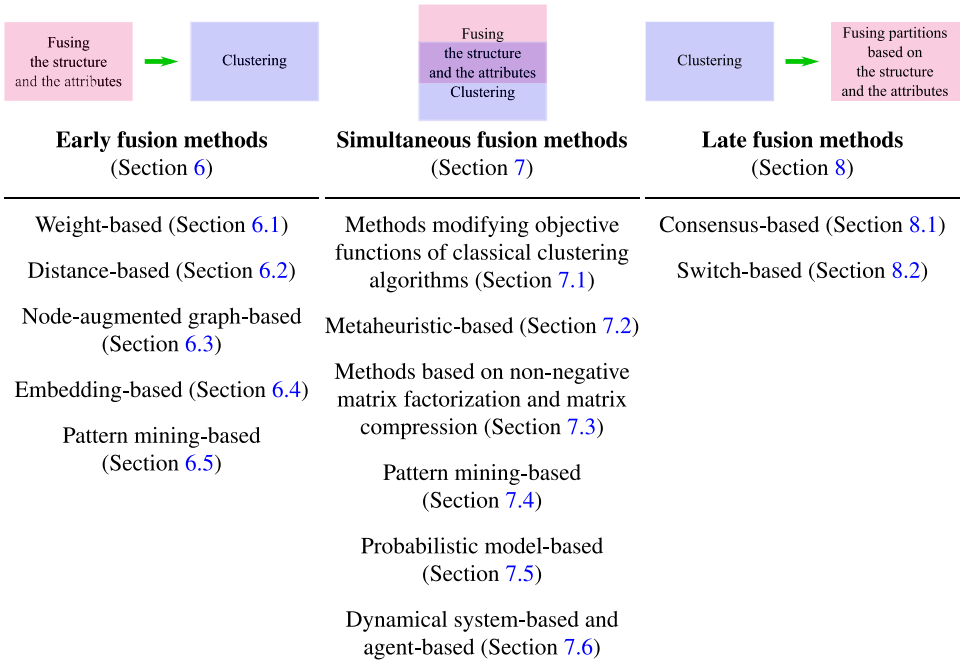

Fig. 2. The proposed classification and the survey structure guide.

provide a list of other methods for *community detection in node-attributed networks* (using both structure and attributes) that the method under consideration is compared with. Note that the list may be empty sometimes. This is so, for example, if the method under consideration is compared only with classical community detection methods that deal either with structure or attributes and do not fuse them to detect communities.

5. Most used node-attributed network datasets and quality measures for community detection evaluation

5.1. Datasets

The title of the survey suggests that we are focused on community detection in node-attributed *social* networks. However, the methods that are included in our classification, generally speaking, may be applied to node-attributed networks of different nature. As we have noticed, authors of the methods implicitly share this point of view and freely use various node-attributed network datasets to evaluate community detection quality.

In this subsection (Tables 1, 2 or 3) we collect and briefly describe the datasets that are popular in the field.[3] Recall that the dataset names written in **bold** in Sections 6–8 refer to the tables in this subsection. It can be observed that Tables 1, 2 or 3 contain datasets based on social network data (e.g. Facebook, LastFM and Twitter) and document or citation network data (e.g. DBLP, Wiki and Patents).

For convenience, we distinguish the datasets by size. Namely, by *small*, *medium* and *large* we mean network datasets with $< 10^3$, $10^3 \dots 10^5$ and $> 10^5$ nodes, correspondingly.

[3] An interested reader can find more node-attributed network datasets e.g. at Mark Newman page, HPI Information Systems Group, LINQS Statistical Relational Learning Group, Stanford Large Network Dataset Collection, University of Verona Laboratory of Cell Trafficking and Signal Transduction, Marc Plantevit page, Tore Opsahl page, UCINET networks, Interactive Scientific Network Data Repository, and Citation Network Dataset.

5.2. Community detection quality measures

Given a set of detected communities (overlapping or not), one needs to evaluate their quality. There are two possible options for this depending on the network dataset under consideration. If the network dataset has no ground truth, one can use various measures of structural closeness and attribute homogeneity. According to our observations, the most popular quality measures in this case are Modularity and Density for the former and Entropy for the latter. Many others such as Conductance, Permanence, Intra- and Inter-Cluster Densities, etc., are also possible. If there is ground truth, it is traditional to compare the detected communities with the ground truth ones. This can be done, for instance, with the following popular measures: Accuracy, Normalized Mutual Information (denoted below by NMI), Adjusted Rand Index or Rand Index (denoted below by ARI and RI, correspondingly) and F-measure. We will discuss both the approaches further in Section 9.

Due to space limitations, we refer the reader to the comprehensive survey [33] and to [24, Sections 2.2 and 4], where all the above-mentioned quality measures and many others are precisely defined and discussed in detail.

6. Early fusion methods

As we have already mentioned, these methods fuse structure and attributes before the community detection process so that the data obtained are suitable for classical community detection algorithms (thus one can use their existing software implementations).

Before we proceed, we introduce additional notation applied only to the weight-based and distance-based early fusion methods. The fact is that existing network structure may be saved or modified depending on the heuristics used in a method, therefore we distinguish

- *fixed topology methods* that use initial network structure without modifying it with respect to attributes,
- *non-fixed topology methods* that modify initial network structure with respect to attributes, in particular, add/erase edges and/or vertices.

Table 4
Weight-based methods with $\alpha = 0$ in (6.1).

Method	Community detection method used and its **input**	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Topology	Datasets used	Compared with
WNev [142]	**Weighted graph** MinCut [105] MajorClust [171] Spectral [103]	No/No	Small	Accuracy	Fixed	Synthetic	–
WSte1 [172]	**Weighted graph** Threshold	No/No	Large	Modularity	Fixed	Phone Network [128]	–
WSte2 [173]	**Similarity matrix (via Weighted graph and random walks)** Hierarchical clustering [66,104]	No/No	Large	Modularity	Fixed	Phone Network [128]	–
WCom1 [42]	**Weighted graph** Weighted Louvain [21]	Yes/No	Small	Accuracy	Fixed	**DBLP***	**WCom2** [42] **DCom** [42]
WCom2 [42]	**Distance matrix (via weighted graph)** Hierarchical agglomerative clustering	Yes/No	Small	Accuracy	Fixed	**DBLP***	**WCom1** [42] **DCom** [42]
AA-Cluster [5,6]	**Node embeddings (via weighted graph)** k-medoids	Yes/No	Small Medium Large	Density Entropy	Fixed	**Political Blogs** **DBLP*** **Patents*** Synthetic	**SA-Cluster** [211] **BAGC** [199] **CPIP** [126]
PWMA-MILP [9]	**Weighted graph** Linear programming MILP [9]	No/No	Small	RI NMI	Fixed	**WebKB**	–
KDComm [19]	**Weighted graph** Iterative Weighted Louvain	No/No	Small Medium Large	F-measure Jaccard Rank Entropy	Fixed	**ego-G+** **Twitter*** **DBLP*** [102] Reddit	**CPIP** [126] **JCDC** [209] **UNCut** [205] **SI** [144]

As far as we know, there is no study on which approach is preferable. How each one influences community detection results is yet to be established.

6.1. Weight-based methods

These methods (see Tables 4 and 5) convert attributes $(\mathscr{V}, \mathscr{A})$ into a weighted *attributive* graph and further somehow merge it with structural graph $(\mathscr{V}, \mathscr{E})$. The result is a weighted graph G_W that is a substitution for node-attributed graph G, see Fig. 3. Edge weights of G_W are usually assigned as follows:

$$W_\alpha(v_i, v_j) = \alpha w_S(v_i, v_j) + (1-\alpha) w_A(v_i, v_j),$$
$$\alpha \in [0, 1], \qquad v_i, v_j \in \mathscr{V}, \tag{6.1}$$

where w_S and w_A are chosen *structural* and *attributive* similarity functions, respectively. The hyperparameter α controls the balance between structure and attributes. Clearly, if $\alpha = 1$ in (6.1), one obtains weights based only on structure; if $\alpha = 0$, then they are based only on attributes. As for w_S and w_A, $w_S(v_i, v_j)$ usually reflects existing connections in $(\mathscr{V}, \mathscr{E})$ (e.g. $w_S(v_i, v_j) = 1$, if $e_{ij} \in \mathscr{E}$, and $w_S(v_i, v_j) = 0$, otherwise), while $w_A(v_i, v_j)$ may be Cosine Similarity or Matching Coefficient values for vectors $A(v_i)$ and $A(v_j)$.

Once the weighted graph G_W is constructed, one can use classical graph clustering algorithms on it such as Weighted Louvain [21]. Sometimes G_W is instead converted to a certain distance matrix that is further used for detecting communities via distance-based clustering algorithms such as k-means or k-medoids.

It is worth mentioning that the fixed topology methods in this subclass assume that the weights (6.1) are assigned on the same set of edges $\mathscr{E}$ as in the initial node-attributed graph G, while the non-fixed topology ones assign weights on all (or on the most part of) possible edges between nodes in $\mathscr{V}$.

Now let us say some words how the hyperparameter α can be chosen. A very popular approach in fixed topology methods is assuming $\alpha = 0$ in (6.1), see Table 4. This actually means that weights in G_W are based only on attributive similarity. Clearly, this may lead to dominance of attributes in the resulting fusion and disappearance of structural connections between nodes with dissimilar attributes. Varying α over the segment $[0, 1]$ in (6.1), used by the methods in Table 5, seems more adequate for controlling the impact of both the components. However, tuning α in this case is usually performed *manually*. In fact, we are unaware of any general non-manual approaches for tuning α in (6.1), although the need for such approaches has been repeatedly emphasized [24].

Note that the choice of similarity functions w_S and w_A is usually determined by preferences of the authors of a particular method. The systematic study of how such a choice influences the community detection results is yet to be done.

Remark 1. The weight-based strategy has been applied to more general networks than considered in the present survey. For example, [17] and [150] use a similar scheme to detect communities in *multi-layer* networks. Another example is SANS [151] that works with *directed* node-attributed graphs. Edge weighting similar to (6.1) is also applied in FocusCO [155]. Although it is not a purely unsupervised community detection method (it requires user's preferences on some of the attributes), it can simultaneously extract local clusters and detect outliers in a node-attributed social network.

Remark 2. Let us mention that there exist community detection methods similar ideologically (attributes → edge weights → weighted graph node embeddings[4] → k-means) but preceding to recently proposed **AA-Cluster** [5,6]. Namely, GraphEncoder [181] and GraRep [31] also first convert a node-attributed graph into a weighted one according to (6.1) with $\alpha = 0$ and then find corresponding node embeddings. The embeddings are further fed to

[4] Low-dimensional continuous vector representations of graph nodes. See also Section 6.4.

Table 5
Weight-based methods with $\alpha \in [0, 1]$ in (6.1).

Method	α in (6.1)	Community detection method used and its **input**	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Topology	Datasets used	Compared with
WWan [190]	[0, 1] in theory 1/2 in experiments	**Edge similarity matrix (via weighted graph)** EdgeCluster [178]	Yes	Small	NMI Micro-F1 Macro-F1	Non-fixed: removing edges	Synthetic BlogCatalog **Delicious***	Non-overlapping co-clustering [55]
SAC2 [52]	[0, 1]	***k*NN (unweighted) graph (via weighted graph)** (Unweighted) Louvain [21]	No/No	Small Medium	Density Entropy	Non-fixed: removing edges	**Political Blogs Facebook100 DBLP10K**	**SAC1** [52] **WSte2** [173] Fast greedy [40] for weighted graph
WCru [48,49]	[0, 1] in theory Not specified in experiments	**Weighted graph** Weighted Louvain [21]	No	Medium	Modularity Intracluster distance	Fixed	**Twitter***	–
CODICIL [160]	[0, 1] in theory 1/2 in some experiments	**Weighted graph** Metis [106] Markov Clustering [165]	No	Small Medium Large	*F*-measure	Non-fixed: adding and removing edges	**CiteSeer*** **Flickr*** Wikipedia*	**Inc-Cluster** [212] **PCL-DC** [202] Link-PLSA-LDA [140]
WMen [132]	Not specified	**Weighted graph/Distance matrix for the weighted graph** SLPA [195] Weighted Louvain [21] K-medoids [207]	Yes–No/ Yes–No	Small Medium	NMI *F*-measure Accuracy	Fixed	**Lazega Research Consult** LFR benchmark [113]	**CODICIL** [160] **SA-Cluster** [211]
PLCA-MILP [9]	[0, 1]	**Weighted graph** Linear programming MILP [9]	No/No	Small	RI NMI	Non-fixed: adding and removing edges	**WebKB**	**SCD** [123] **ASCD** [159] **SCI** [187] PCL-DC [202] Block-LDA [14]
kNN-enhance [102]	May be thought 1/2 (*k*NN by attributes)	**Distance matrix (of an edge-augmented graph)** *k*NN *k*-means	No/No	Medium	Accuracy NMI F-Measure Modularity Entropy	Non-fixed: adding edges	**Cora Citeseer Sinanet PubMed Diabetes DBLP***	**PCL-DC** [202] PPL-DC [201] **PPSB-DC** [32] **CESNA** [204] **cohsMix** [208] **BAGC** [198] **GBAGC** [199] **SA-Custer** [211] **Inc-Cluster** [212] **CODICIL** [160] GLFM [125]
IGC-CSM [141] (source)	[0, 1] in theory 1/2 in experiments	**Distance matrix for the weighted graph** *k*-Medoids	Yes/ No	Medium	Density Entropy	Fixed	**Political Blogs DBLP10K**	**SA-Cluster** [211] **SA-Cluster-Opt** [39]
AGPFC [90]	[0, 1] in theory Manually tuned in experiments	Fuzzy equivalent matrix λ-cut set method	No/Yes	Small Medium	Density Entropy	Fixed	**Political Blogs CiteSeer Cora WebKB**	**SA-Cluster** [211] **BAGC** [198]
NMLPA [96]	1/2	**Weighted graph** A multi-label propagation algorithm	Yes/ Yes	Medium	*F*1-score Jaccard	Fixed	**ego-Facebook Flickr*** [160] **ego-Twitter**	**CESNA** [204] **SCI** [187] **CDE** [124]

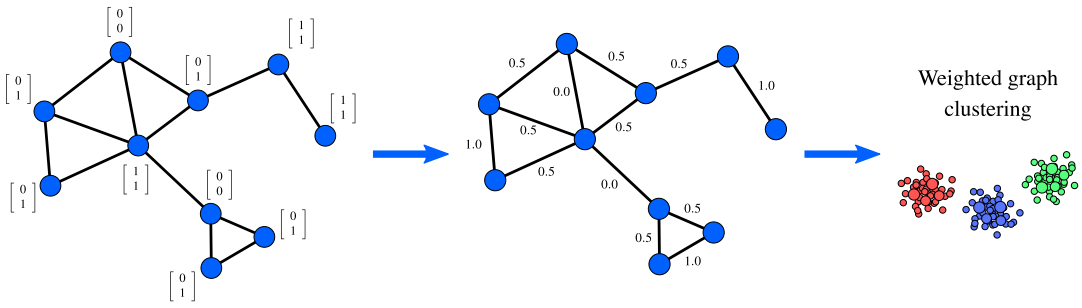

Fig. 3. A typical scheme of a weight-based method (the attributive weights here are the values of normalized matching coefficient for the attribute vectors).

k-means algorithm to detect communities. However, in opposite to [5,6,181] and [31] mostly focus on node embedding techniques for a weighted graph than on the community detection task.

6.2. Distance-based methods

Methods discussed in the previous subsection aim at representing structure and attributes in a unified graph form suitable for further graph clustering. In opposite, distance-based methods intentionally abandon graph representation in favor of the representation via a distance matrix that contains information about structure and attributes. The distance matrix is usually obtained by fusing the components by a structure- and attributes-aware distance function, see Fig. 4. The matrix can be further fed to distance-based clustering algorithms such as k-means and k-medoids. Note that the resulting clusters may contain disconnected portions of initial graph G as the graph structure is removed at the fusion step [5, Section 3.3].

The distance function fusing structure and attributes is often of the form

$$D_\alpha(v_i, v_j) = \alpha d_S(v_i, v_j) + (1-\alpha) d_A(v_i, v_j), \quad \alpha \in [0, 1], \qquad v_i, v_j \in \mathscr{V}, \tag{6.2}$$

where d_S and d_A are *structural* and *attributive* distance functions, correspondingly. As in the case of (6.1), α in (6.2) controls the balance between the components; how to choose it "properly" seems to be an open problem, too. As for the distance functions, it is common to define $d_S(v_i, v_j)$ as the shortest path length between v_i and v_j. Possible options for $d_S(v_i, v_j)$ are the Jaccard or Minkowski distances between vectors $A(v_i)$ and $A(v_j)$.

Short descriptions for known distance-based methods are given in Table 6. Note that **ANCA** [61,62] and **SToC** [15] employ a bit different fusion than in (6.2) but nevertheless still deal with certain structure- and attribute-aware distances.

Remark 3. There exist distance-based methods for *multi-layer* networks. For example, CLAMP [149] is an method for clustering networks with heterogeneous attributes and multiple types of edges that uses a distance function similar to (6.2), in a sense.

6.3. Node-augmented graph-based methods

Methods from this subsection (see Table 7) transform initial node-attributed graph G into another node-augmented graph $\tilde{G}$, with nodes from $\mathscr{V}$ and new *attributive* nodes representing attributes, see Fig. 5. To be more precise, suppose that $dom(a_k) = \{s_l\}_{l=1}^{L_k}$, i.e. the domain of kth element (attribute) of attribute vector A contains L_k possible values. Then one should create L_k new attribute nodes $\tilde{v}_{k,l}$ corresponding to lth value of kth attribute. Such a procedure is performed for $k = 1, \ldots, d$, where $d = \dim A$. The set $\mathscr{V}_A := \{\tilde{v}_{k,l}\}$, where $k = 1, \ldots, d$ and $l = 1, \ldots, L_k$, is then the set of attributive nodes. An edge between structural node $v_i \in \mathscr{V}$ and attributive node $\tilde{v}_{k,l}$ in $\tilde{G}$ exists if $a_k(v_i) = s_l$. Community detection is further performed in $\tilde{G}$. The methods in Table 7 propose to apply random walks [183] to obtain a certain distance matrix for G and further use it in a distance-based clustering algorithm.

Note that the above-mentioned augmentation is not applicable to continuous attributes. What is more, $\tilde{G}$ contains much more nodes end edges than G (especially if d and L_k, $k = 1, \ldots, d$, are large) and this makes the methods from this subclass rather computationally expensive.

6.4. Embedding-based methods

As is well-known, a graph as a traditional representation of a network brings several difficulties to network analysis. As mentioned in [51], graph algorithms suffer from high computational complexity, low parallelisability and inapplicability of machine learning methods. Novel embedding techniques aim at tackling this by learning *node embeddings*, i.e. low-dimensional continuous vector representations for network nodes so that main network information is efficiently encoded [29,51,74]. Roughly speaking, node embedding techniques allow to convert a graph with n nodes into a set of n vectors.

In the context of node-attributed social networks, the objective of node embedding techniques is efficient encoding both structure and attributes [31,68,179]. We are not going to provide general details on the techniques as this has been already done in the surveys [29,51]. Let us just mention that having node embeddings (i.e. vectors) at hand allows one to use classical distance-based clustering algorithms such as k-means to detect communities, see Fig. 6.

It turns out that there exists a rich bibliography on node embedding techniques for attributes networks of different type [29, 51]. However, not all of them have been applied to the community detection task (the classification task is typically considered). Taking this into account, we confine ourselves in this survey only to the embedding-based (early fusion) methods that have been used for community detection in node-attributed social or document networks and compared with other community detection methods. The results are presented in Table 8.

Remark 4. Various embedding techniques applicable for community detection in *multi-layer* networks are considered e.g. in [36, 95,154].

6.5. Pattern mining-based (early fusion) methods

Recall that a motif is a pattern of the interconnection occurring in real-world networks at numbers that are significantly higher than those in random networks [133]. Motifs are considered as building blocks for complex networks [133]. We found just one community detection method for node-attributed social networks using such patterns, namely, **AHMotif** [120], see Table 9. This method equips structural motifs identified in the network with the so-called homogeneity value based on node attributes involved in the motif. This information is stored in a special adjacency matrix that can be an input to classical community detection algorithms.

7. Simultaneous fusion methods

Recall that simultaneous fusion methods fuse structure and attributes in a joint process with community detection. For this reason, these methods often require special software implementation, in contrast to early and late fusion methods that partially allow one to use existing implementations of classical community detection algorithms.

7.1. Methods modifying objective functions of classical clustering algorithms

Table 10 contains[5] short descriptions of simultaneous fusion methods that modify objective functions of well-known clustering algorithms such as Louvain, Normalized Cut, k-means, k-medoids and kNN. Their main idea is to adapt a classical method

[5] The authors of **ILouvain** [43] claim that they compare their method with ToTeM [42], "another community detection method designed for attributed graphs". However, it seems that there is an inaccuracy with it as we could not find in [42] any method called ToTeM.

Table 6
Distance-based methods.

Method	α in (6.2)	Community detection method used and its **input**	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Topology	Datasets used	Compared with
DCom [42]	[0, 1]	**Distance matrix** Hierarchical agglomerative clustering	Yes/No	Small	Accuracy	Non-fixed: adding edges	**DBLP***	**WCom1** [42] **WCom2** [42]
DVil [148,186]	[0, 1]	**Distance matrix** Self-organizing maps [148,186]	No/No	Small Medium	NMI	Non-fixed: adding edges	Synthetic Medieval Notarial Deeds	–
SToC [15]	May be thought depending on the values of d_S and d_A	**Distance matrix** τ-close clustering [15]	No/No	Medium Large	Modularity Within-Cluster Sum of Squares	Non-fixed: adding edges	**DBLP10K** DIRECTORS* DIRECTORS-gcc*	**Inc-Cluster** [212] **GBAGC** [199]
@NetGA [156]	[0, 1] in theory 1/2 in experiments	**Distance matrix** Genetic algorithm	No/No	Medium	NMI	Non-fixed: adding edges	Synthetic	**SA-Cluster** [211] CSPA [58,174] **Selection** [58]
ANCA [61,62]	May be thought 1/2	**Distance matrix** k-means	Yes/No	Medium	ARI NMI Density Modularity Conductance Entropy	Fixed	Synthetic **DBLP10K** Enron email corpus	**SA-Cluster** [211] **SAC1–SAC2** [52] **IGC-CSM** [141] **WSte1** [172] **ILouvain** [43]

Table 7
Node-augmented graph distance-based methods.

Method	Graph augmentation	Community detection method used and its **input**	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
SA-Cluster [211] **Inc-Cluster** [38,212] **SA-Cluster-Opt** [39]	New nodes and edges	**Distance matrix (via neighborhood random walks)** Modified k-medoids [211]	Yes/No	Small Medium	Density Entropy	**Political Blogs** **DBLP10K** **DBLP84K**	W-Cluster [211] (based on (6.2)) **SA-Cluster** [211] **Inc-Cluster** [38,212] **SA-Cluster-Opt**
SCMAG [98]	New nodes and edges	**Distance matrix (via neighborhood random walks)** Subspace clustering algorithm based on ENCLUS [37]	No/Yes	Medium	Density Entropy	IMDB Arnetminer bibliography*	**SA-Custer** [211] GAMer [83]

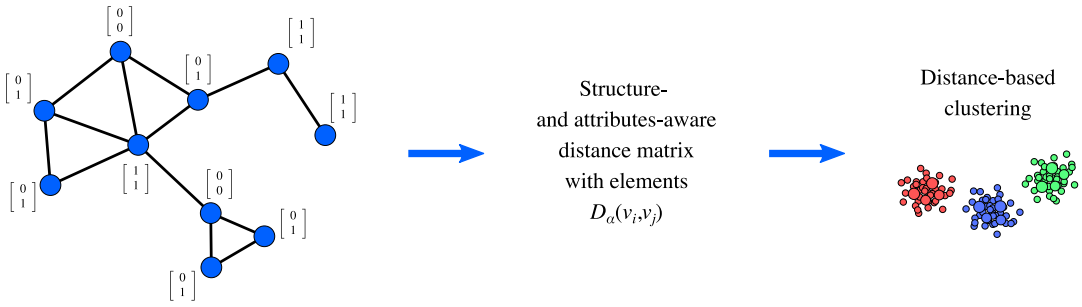

Fig. 4. A typical scheme of a distance-based method.

(that works, for example, only with network structure originally) for using both structure and attributes in the optimization process. For example, if one wants to modify Louvain [21] whose original objective function is structure-aware Modularity, one can include an attributes-aware objective function, say, Entropy in the optimization process. Then Modularity is maximized and Entropy is minimized simultaneously in an iterative process similar to that of Louvain [21].

7.2. Metaheuristic-based methods

Methods in this subclass are rather similar ideologically to those in Section 7.1. However, instead of modifying objective functions and iterative processes of well-known community detection algorithms, they directly apply metaheuristic algorithms (in particular, evolutionary algorithms and tabu search) to find a node-attributed network partition that provides optimal values for certain measures of structural closeness and attribute homogeneity. More precisely, they use metaheuristics for optimization of a combination of structure- and attributes-aware

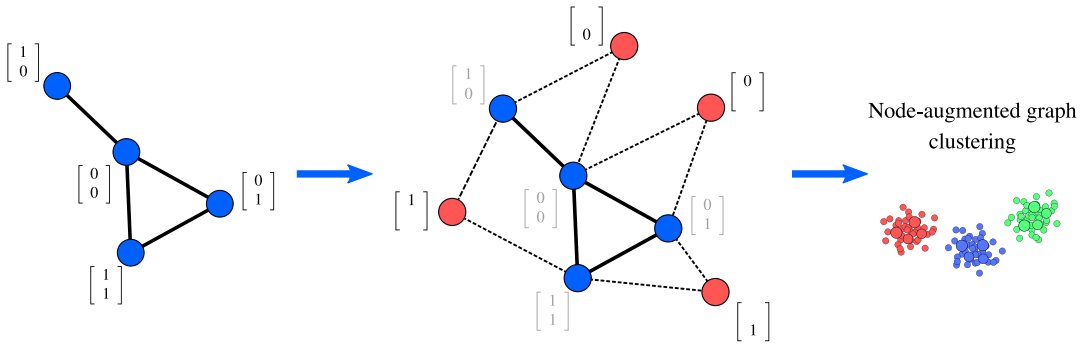

Fig. 5. A typical scheme of a node-augmented graph-based method.

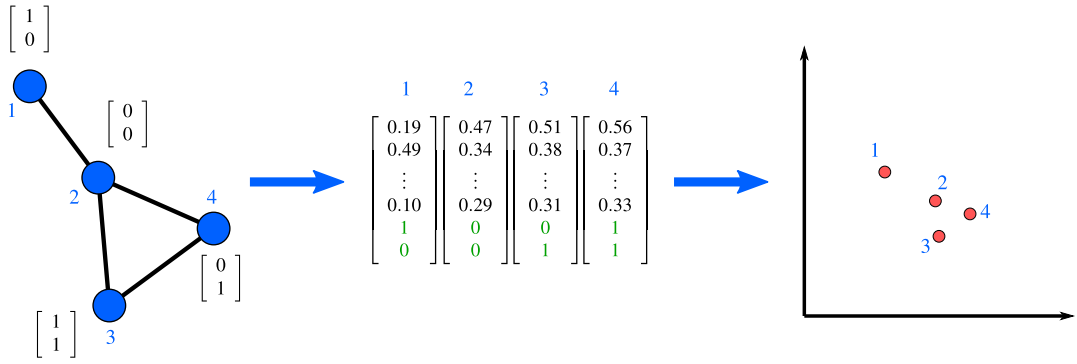

Fig. 6. A possible scheme of an embedding-based method (the attribute vectors are here concatenated with the node embeddings and then fed to k-means).

Table 8
Embedding-based methods.

Method	Embedding technique	Community detection method used and its **input**	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
PLANE [115]	A generative model and EM [53]	**Node embeddings** k-means	Yes/No	Small Medium	Accuracy	**Cora***	Relational Topic Model [35]+Topic Distributions Embedding [101]
DANE [68]	Autoencoder	**Node Embeddings** k-means	Yes/No	Medium	Accuracy	**Cora** **Citeseer** **PubMed Diabetes** **Wiki**	Embeddings obtained via TADW [203] LANE [94] GAE [108] VGAE [108] GraphSAGE [86]
CDE [124]	Structure embedding matrix	**Structure embedding matrix and attribute matrix** Non-negative matrix factorization	Yes/Yes-No	Small Medium	Accuracy NMI Jaccard F1-score	**Cora** **Citeseer** **WebKB** **Flickr*** **ego-Facebook** Philosophers [4]	**PCL-DC** [202] **Circles** [118] **CESNA** [204] **SCI** [187]
MGAE [189]	Autoencoder	**Node embeddings** Spectral clustering	Yes/No	Medium	Accuracy NMI F-score Precision Recall Average Entropy ARI	**Cora** **CiteSeer** **Wiki**	**Circles** [118] RTM [35] RMSC [194] Embeddings obtained via TADW [203] VGAE [108]

Table 9
Pattern mining-based (early fusion) methods.

Method	Pattern used	Community detection method used and its **input**	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
AHMotif [120]	Motif	**Structure- and attributes-aware adjacency matrix** Permanence [34] Affinity Propagation [67]	Yes/No	Medium	NMI Accuracy	**Cora** **WebKB**	–

Table 10
Methods modifying objective functions of classical clustering algorithms.

Method	Modified algorithm	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
OCru [49]	Louvain [21] Added attribute Entropy minimization	No/No	Medium	Modularity Entropy	**Facebook100**	–
SAC1 [52]	Louvain [21] Added attribute similarity maximization	No/ No	Small Medium	Density Entropy	**Political Blogs** **Facebook100** **DBLP10K**	**SAC2** [52] **WSte2** [173] Fast greedy [40] for weighted graph
ILouvain [43] (source)	Louvain [21] Added maximization of attribute-aware Inertia	No/ No	Small Medium	NMI Accuracy	DBLP+Microsoft Academic Search* Synthetic	ToTeM [42]
LAA/LOA [11]	Louvain [21] Modularity gain depends on attributes	No/No	Small	Density Modularity	London gang [75] Italy gang Polbooks Adjnoun [143] Football [72]	–
MAM [163] (source)	Louvain-type algorithm with attribute-aware Modularity+Outlier detection	No/No	Small Medium Large	F1-score Attribute-aware Modularity	Synthetic Disney [137] DFB [84] ARXIV [84] IMDB [84] **DBLP*** **Patents*** Amazon [164]	CODA [69]
UNCut [205]	Normalized Cut Added attributes-aware Unimodality Compactness	Yes/No	Small Medium	NMI ARI	Disney [137] DFB [84] ARXIV [84] **Political Blogs** DBLP-4AREA [155] **Patents**	**SA-cluster** [211] SSCG [84] **NNM** [169]
NetScan [60,70]	An approximation algorithm for the connected *k*-Center optimization problem (structure and attributes involved)	Yes/Yes	Small Medium	Accuracy	Professors* Synthetic **DBLP*** BioGRID+Spellman	–
JointClust [136]	An approximation algorithm for the Connected X Clusters problem (structure and attributes involved)	No/No	Medium	Accuracy	**DBLP*** **CiteSeer*** Corel stock photo collection	–
SS-Cluster [63]	*k*-Medoid with structure- and attributes-aware objective functions	Yes/No	Medium	Density Entropy	**Political Blogs** **DBLP10K**	**SA-cluster** [39,211] W-cluster [39] *k*SNAP [182]
Adapt-SA [121]	Weighted *k*-means for *d*-dimensional representations of structure and attributes	Yes/No	Medium	Accuracy NMI F-measure Modularity Entropy	Synthetic **WebKB** **Cora** **Political Blogs** **CiteSeer** **DBLP10K**	**CODICIL** [160] **SA-Cluster** [213] **Inc-Cluster** [212] **PPSB-DC** [32] **PCL-DC** [202] **BAGC** [198]
kNAS [22]	*kNN* with added Semantic Similarity Score	Yes/Yes	Medium	Density Tanimoto Coefficient	**DBLP*** **Facebook*** **Twitter***	**SA-Cluster-Opt** [39] **CODICIL** [160] NISE [192]

objective functions, for example, Modularity and Attributes Similarity. Short descriptions of the methods from this subclass are given in Table 11.

7.3. Methods based on non-negative matrix factorization and matrix compression

Non-negative matrix factorization (NNMF) is a matrix technique that consists in approximating a non-negative matrix with high rank by a product of non-negative matrices with lower ranks so that the approximation error by means of the Frobenius norm[6] F is minimal. As is well known, NNMF is able to find clusters in the input data [116].

To be applied to community detection in node-attributed social networks, NNMF requires a proper adaptation to fuse both structure and attributes. Different versions of such an adaptation have been proposed, see Table 12.

To be more formal in describing the corresponding NNMF-based methods, let us introduce additional notation. Let $\mathbf{S}_{n\times n}$ denote the adjacency matrix for the network structure (as before, n is the number of nodes), $\mathbf{A}_{n\times d}$ the node attribute matrix for the network attributes (d is the dimension of attribute vector A), N the number of required clusters (it is an input in NNMF-based methods), $\mathbf{U}_{n\times N}$ the cluster membership matrix whose elements indicate the association of nodes with communities,

[6] A matrix norm defined as the square root of the sum of the absolute squares of matrix elements.

Table 11
Metaheuristic-based methods.

Method	Metaheuristic optimization algorithm	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
MOEA-SA [122]	Multiobjective evolutionary algorithm (Modularity and Attribute Similarity are maximized)	No/No	Small Medium	Density Entropy	**Political Books** **Political Blogs** **Facebook100** **ego-Facebook**	**SAC1–SAC2** [52] **SA-Cluster** [211]
MOGA-@Net [157]	Multiobjective genetic algorithm (optimizing Modularity, Community score, Conductance, attribute similarity)	No/No	Small Medium	NMI Cumulative NMI Density Entropy	Synthetic **Cora** **Citeseer** **Political books** **Political Blogs** **ego-Facebook**	**SA-cluster** [211] **BAGC** [198] **OCru** [50] **Selection** [58] HGPA-CSPA [58,174]
JCDC [209]	Tabu search and gradient ascent for a structure- and attributes-aware objective function	Yes/No	Small Medium	NMI	Synthetic World trade [147] **Lazega**	CASC [20] **CESNA** [204] **BAGS** [198]

and finally $\mathbf{V}_{d\times N}$ denotes the cluster membership matrix whose elements indicate the association of the attributes with the communities. In these terms, the aim of NNMF-based methods is to use known matrices **S**, **A** and the number of clusters N in order to determine the unknown matrices **U** and **V** in an iterative optimization procedure. For example, **SCI** [187] models structural closeness as $\min_{\mathbf{U}\geqslant 0}\|\mathbf{S}-\mathbf{U}\mathbf{U}^T\|_F^2$ and attribute homogeneity as $\min_{\mathbf{V}\geqslant 0}\|\mathbf{U}-\mathbf{A}\mathbf{V}\|_F^2$. It is also proposed to select the most relevant attributes for each community by adding an l_1 norm sparsity term to each column of matrix **V**. As a result, one obtains the following optimization problem:

$$\min_{\mathbf{U}\geqslant 0,\,\mathbf{V}\geqslant 0}\left(\alpha_1\|\mathbf{S}-\mathbf{U}\mathbf{U}^T\|_F^2+\|\mathbf{U}-\mathbf{A}\mathbf{V}\|_F^2+\alpha_2\sum_j\|\mathbf{V}(\cdot,j)\|_1^2\right),$$

where $\alpha_1>0$ controls the impact of structure and $\alpha_1\geqslant 0$ the sparsity penalty. This problem is further approximately solved in an iterative process according to Majorization–Minimization framework [4].

Remark 5. An NNMF-based community detection method for *multi-layer* networks is given in [100].

Now we briefly describe **PICS** [7], the method adopting matrix compression[7] for community detection in node-attributed social networks. **PICS** uses lossless compression principles from [76] to simultaneously compress the network adjacency matrix **S** and the attribute matrix **A**. As a result of the compression, certain homogeneous rectangular blocks in the matrices can be determined. Groups of the nodes corresponding to the blocks are considered as communities. One should be aware however that nodes within communities found by **PICS** may not be densely connected due to the definition of a community in [7].

7.4. Pattern mining-based (simultaneous fusion) methods

Pattern mining in node-attributed social networks focuses on extraction of patterns, e.g. subsets of specific attributes or connections,[8] in network structure and attributes [13]. Among other things, this helps to make sense of a network and to understand how it was formed. In the context of community detection, the extracted patterns are used as building blocks for communities.

There are many papers devoted to pattern mining in social networks [13] but the majority of them do not deal with the task of community detection. The ones relevant to the topic of this survey are presented in Table 13. It is worth mentioning here that it is common for pattern mining-based methods to detect communities not in the whole network but in its part only (e.g. [12,158]).

Remark 6. ABACUS [18] detects communities by extracting patterns in *multi-layer* networks.

7.5. Probabilistic model-based methods

Methods from this subclass probabilistically infer the distribution of community memberships for nodes in a node-attributed social network under the assumption that network structure and attributes are generated according to chosen parametric distributions. Generative and discriminative models are mainly used for the inferring. It is worth mentioning that it is though a non-trivial task to "properly" choose a priori distributions for structure and attributes [5].

Short descriptions of the methods from this subclass are given in Table 14. Pay attention that this table does not contain any method preceding to [202] and this requires the following additional comments. According to [202], several probabilistic model-based clustering methods for node-attributed networks had been proposed before [202], for example, in [41,59,140]. However, they focus on node-attributed *document* networks which are out of scope of the present survey. That is why they are non included in Table 14.

Remark 7. TUCM [161] proposes a generative Bayesian model for detecting communities in *multi-layer* networks where different types of interactions between social actors are possible.

7.6. Dynamical system-based and agent-based methods

Methods from this subclass (see Table 15) treat a node-attributed social network as a dynamic system and assume that its community structure is a consequence of certain interactions among nodes (of course, the attributes are thought to affect the interactions). Some of the methods assume that the interactions occur in an information propagation process, i.e. while information is sent to or received from every node. Others comprehend each node as an autonomous agent and develop a multi-agent system to detect communities. Note that these approaches are rather recent and consider community detection in node-attributed social networks from a new perspective. Furthermore, they seem to be efficient for large networks as can be easily parallelized.

[7] The aim of compression methods is to find a shorter form of describing the same information content.

[8] An example of a pattern is a maximal clique [24,107]. Recall that a clique is a subset of nodes in an undirected graph such that every two nodes are adjacent. A clique is called maximal if there is no other clique that contains it.

Table 12
Methods based on non-negative matrix factorization and matrix compression.

Algorithm	Factorization/ compression type	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
NPei [153]	3-factor NNMF	Yes/Yes	Small Medium	Purity	**Twitter** **DBLP***	Relational Topic Model [35]
3NCD [146]	2-factor NNMF	Yes/Yes	Medium Large	F1-score Jaccard	**ego-Facebook** **ego-Twitter** **ego-G+**	**CESNA** [204]
SCI [187]	2-factor NNMF	Yes/Yes	Medium	Accuracy NMI GNMI *F*-measure Jaccard	**Citeseer** **Cora** **WebKB** **LastFM**	**PCL-DC** [202] **CESNA** [204] **DCM** [158]
JWNMF [98]	2-factor NNMF	Yes/Yes	Small Medium	Modularity Entropy NMI	Amazon Fail Disney Enron DBLP-4AREA **WebKB** **Citeseer** **Cora**	**BAGC** [198] **PICS** [7] SANS [151]
SCD [123]	2- and 3-factor NNMF	Yes/Yes–No	Small Medium	Accuracy NMI	**Twitter** **WebKB**	**SCI** [187]
ASCD [159]	2-factor NNMF	Yes/Yes–No	Small Medium	Accuracy NMI *F*-measure Jaccard	**LastFM** **WebKB** **Cora** **Citeseer** **ego-Twitter*** **ego-Facebook***	Block-LDA [14] **PCL-DC** [202] **SCI** [187] **CESNA** [204] **Circles** [130]
CFOND [85]	2- and 3-factor NNMF	Yes/Yes-No	Medium	Accuracy NMI	**Cora** **CiteSeer** **PubMed** **Attack** Synthetic	GNMF [28] DRCC [77] LP-NMTF [188] iTopicModel [175]
MVCNMF [88]	2-factor NNMF	Yes/Yes	Small Medium	Density Entropy	**Political Blogs** **CiteSeer** **Cora** **WebKB** ICDM-DBLP	FCAN [93] SACTL [197] **kNAS** [22]
PICS [7] (source)	Matrix compression (finding rectangular blocks)	No/No	Small Medium	Anecdotal and visual study	Youtube [134] **Twitter*** Phonecall [57] Device [57] **Political Books** **Political Blogs**	–

Table 13
Pattern mining-based (simultaneous fusion) methods.

Method	Patterns used	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
DCM [158] (source)	Semantic patterns (*queries*)	Yes/Yes	Small Medium	Community score Conductance Intra-cluster density Modularity	**Delicious** **LastFM** **Flickr**	–
COMODO [12]	Semantic patterns	Yes/Yes	Small Medium	Description complexity Community size	BibSonomy [16] **Delicious** **LastFM**	**DCM** [158]
ACDC [107]	Maximal cliques	Yes/Yes	Medium	Density	**Political Blogs**	**SA-Cluster** [211] **SAC1–SAC2** [52]

8. Late fusion methods

Recall that late fusion methods intend to fuse structural and attributive information after the community detection process. More precisely, community detection is first separately performed for structure (e.g. by Louvain [21]) and attributes (e.g. by *k*-means [87]). After that, the partitions obtained are fused somehow in order to get the resulting structure- and attributes-aware partition, see Fig. 7.

Note that late fusion methods usually allow a researcher/data scientist to use existing implementations of classical community detection and consensus clustering algorithms to get the required partition.

8.1. Consensus-based methods

Given a set (also known as an ensemble) of partitions, the general goal of consensus clustering algorithms is to find a single

Table 14
Probabilistic model-based methods.

Method	Model features	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
PCL-DC [202]	Conditional Link Model Discriminative Content model	Yes/No	Medium	NMI Pairwise *F*-measure Modularity Normalized cut	**Cora** **Siteseer**	PHITS-PLSA [41] LDA-Link-Word [59] Link-Content-Factorization [214]
CohsMix [208]	MixNet model [170]	Yes/No	Small	RI	Synthetic Exalead.com search engine dataset	Multiple view learning [210] Hidden Markov Random Field [10]
BAGC [198] **GBAGC** [199]	a Bayesian treatment on distribution parameters	Yes/No	Medium	Modularity Entropy	**Political Blogs** **DBLP10K** **DBLP84K**	**Inc-Cluster** [212] **PICS** [7]
VEM-BAGC [30]	Based on **BAGC** [198]	Yes/No	Medium	Modularity Entropy	**Political Blogs** Synthetic networks	**BAGC** [198]
PPSB-DC [32]	Popularity–productivity stochastic block model and discriminative content model	Yes/No	Medium	NMI Pairwise F-measure Accuracy	**Cora** **CiteSeer** **WebKB**	**PCL-DC** [202] PPL-DC [201]
CESNA [204]	A probabilistic generative model assuming that communities generate network structure and attributes	No/Yes	Medium Large	Evaluation	**ego-Facebook** **ego-G+** **ego-Twitter** **Wikipedia*** **(philosophers)** **Flickr**	**CODICIL** [160] **Circles** [130] Block-LDA [14]
Circles [130]	A generative model for friendships in social circles	Yes/Yes	Medium Large	Balanced Error Rate	**ego-Facebook** **ego-G+** **ego-Twitter**	Block-LDA [14] Adapted Low-Rank Embedding [206]
SI [144]	A modified version of a stochastic block model [91]	Yes/No	Small Medium	NMI	Synthetic High school friendship Food web of marine species in the Weddell Sea Harvard Facebook Malaria HVR 5 and 6 gene recombination	–
NEMBP [89]	A generative model with learning method using a nested EM algorithm with belief propagation	Yes/Yes-No	Small Medium	Accuracy NMI GNMI F-score Jaccard	**WebKB** **ego-Twitter*** **ego-Facebook*** **CiteSeer** **Cora** **Wikipedia*** **Pubmed**	Block-LDA [14] **PCL-DC** [202] **CESNA** [204] **DCM** [158] **SCI** [187]
NBAGC-FABAGC [196]	A nonparametric and asymptotic Bayesian model selection method based on **BAGC** [198]	No/No	Medium	NMI Modularity Entropy	Synthetic **Political Blogs** **DBLP10K** **DBLP84K**	**PICS** [7]

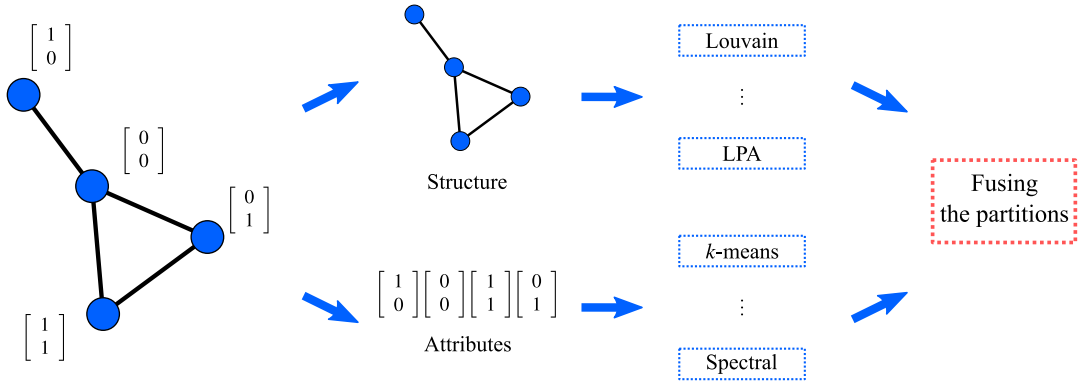

Fig. 7. A typical scheme of a late fusion method.

Table 15
Dynamical system-based and agent-based methods.

Method	Description	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
CPIP-CPRW [126]	Content (information) propagation models: a linear approximate model of influence propagation (**CPIP**) and content propagation with the random walk principle (**CPRW**)	Yes/Yes	Medium	F-score Jaccard NMI	**CiteSeer** **Cora** **ego-Facebook** **PubMed Diabetes**	Adamic Adar [1] **PCL-DC** [202] **Circles** [118] **CODICIL** [160] **CESNA** [204]
CAMAS [26]	Each node with attributes as an autonomous agent with influence in a cluster-aware multiagent system	No/Yes	Medium Large	Coverage Rate Normalized Tightness Normalized Homogeneity F1-Score Jaccard ARI	Synthetic **ego-Facebook** **ego-Twitter*** **ego-G+**	**CESNA** [204] EDCAR [80]
SLA [27]	A dynamic cluster formation game played by all nodes and clusters in a discrete-time dynamical system	Yes/No	Medium Large	Density Entropy F1-score	**Delicious** **LastFM** **ego-Facebook** **ego-Twitter*** **ego-G+**	**CESNA** [204] EDCAR [80]

consolidated partition that aggregates information in the ensemble [78,111,174,176,177]. A recent survey on such methods can be found e.g. in [23]. The idea behind consensus clustering is clearly appropriate for community detection in node-attributed social networks if one has an ensemble of partitions obtained separately (or maybe even jointly) for structure and attributes. Table 16 contains short descriptions of the methods applying the idea.

Remark 8. General-purpose consensus clustering algorithms for *multi-layer* networks are considered in [78,176,177].

8.2. Switch-based methods

The only method included in this subclass (see Table 17) also deals with partitions obtained separately for structure and attributes but chooses a more "preferable" one instead of finding consensus. Namely, **Selection** [58] switches from a structure-based to an attributes-based partition when the former one is *ambiguous*. This refers to the case when the so-called *estimated mixing parameter* μ for the structure-based partition is less then a certain experimental value μ_{lim} associated with a significant drop in clustering quality on synthetic networks [113]. An interested reader can find the precise definitions of μ and μ_{lim} in [58,113].

9. Analysis of the overall situation in the field

The information collected in Sections 6–8 allows us to analyze the overall situation in the field. Particularly, we aim at determining the *state-of-the-art*[9] methods. This is probably what a researcher/data scientist facing the community detection problem in node-attributed social networks expects from our survey.

We start with observing the directed graph based on the information in Sections 6–8 and showing method–method comparisons, see Fig. 8. It requires several comments though. First, there are methods in Sections 6–8 that are not compared with others for community detection in node-attributed networks or compared with a few. For this reason, we include in the graph only nodes (representing methods) whose degree is at least two. This means that there are at least two comparison experiments with each method presented in Fig. 8. Note that 46 methods are shown in the graph of 75 classified in the survey. Second, the directed edges in the graph show the existing method–method comparisons. For example, the directed edge from node **CESNA** [204] to node **CODICIL** [160] indicates that the authors of **CESNA** [204] compared their method with **CODICIL** [160] and showed that **CESNA** [204] outperforms **CODICIL** [160] in some sense (community detection quality, computational efficiency, etc.). This is applied to all edges in the graph.

What is more, we used PageRank to detect the most important or, better to say, *most influential methods* in the field. Nodes with the highest PageRank values are filled green in Fig. 8 so that the darker green means the higher PageRank. It turns out that the most influential ones are (in the order they discussed in Sections 6–8):

- weight-based **SAC2** [52] and **CODICIL** [160] (Section 6.1),
- node-augmented graph-based **SA-Cluster** [211], **Inc-Cluster** [38,212] and **SA-Cluster-Opt** [39] (Section 6.3),
- **SAC1** [52] modifying the Louvain objective function (Section 7.1),
- NNMF-based **SCI** [187] and matrix compression-based **PICS** [7] (Section 7.3),
- pattern mining-based (simultaneous fusion) **DCM** [158] (Section 7.4),
- probabilistic model-based **PCL-DC** [202], **BAGC** [198], **GBAGC** [199], **CESNA** [204] and **Circles** [130] (Section 7.5).

In our opinion, these methods may be though to be chosen by researchers' community as those determining further developments in community detection methods for node-attributed social networks. Thus we encourage a newcomer in the field to get familiar with them first to see the main ideas and techniques existing for community detection in node-attributed social networks. At the same time, we would not consider the most influential methods as state-of-the-art as other methods are shown to outperform them in some sense.

What else the graph in Fig. 8 can tell us about? We emphasize that it does not contain 29 methods of those discussed in Sections 6–8 and furthermore it is rather sparse and even disconnected. These points lead to the conclusion that the comparison study of the methods in the field is far from being complete. We would even strengthen the conclusion made by saying that a researcher/data scientist cannot be sure that the method chosen

[9] According to Cambridge Dictionary, state-of-the-art means "the best and most modern of its type".

Table 16
Consensus-based methods.

Method	Fusing the partitions	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
LCru [47]	Row-manipulation in the contingency matrix for the partitions	No/No	Small Medium	ARI Density Entropy	Facebook* **DBLP10K**	–
Multiplex [97]	Multiplex representation scheme (attributes and structure are clustered separately as layers and then combined via consensus [180])	No/Yes	Medium Large	F1-score	Synthetic **ego-Twitter** **ego-Facebook** **ego-G+**	**CESNA** [204] **3NCD** [146]
WCMFA [127]	Association matrix with weighting based on structure- and attributes-aware similarity	Depends on the partitions	Small	RI ARI NMI	**Consult** [46] London Gang [75] Montreal Gang [54]	**WMen** [132]

Table 17
Switch-based methods.

Method	Fusing the partitions	Require the number of clusters/ Clusters overlap	Size of datasets used for evaluation	Quality measures	Datasets used	Compared with
Selection [58]	Switching between the partitions	Depends on the partitions	Medium	NMI Modularity	Synthetic LFR benchmark [113] **DBLP84K**	**BAGC** [198] **OCru** [49] **SA-Cluster** [211] HGPA-CSPA [174]

for practical use is preferable to other existing ones, even if there is an edge in Fig. 8. The problem is that the comparison experiments (represented via the edges) are made by different means, e.g. different quality measures, datasets, hyperparameter tuning strategies, etc. Let us discuss it below in more detail.

Suppose for a second that two methods show the same community detection quality for a number of datasets and measures, and their hyperparameters are tuned to provide the best possible results. Then we should think how much time/space each method uses for it. Particularly, we may think of method's computational complexity in terms of the number of vertices n, the number of edges m and the attribute dimension d in a node-attributed graph. Such estimates exist for certain methods, particularly for some of the most influential ones. Examples are **CODICIL** [160] with $O(n^2 \log n)$, **SA-Cluster** [211] with $O(n^3)$ and **CESNA** [204] with $O(m)$. However, we could not find such estimates for the majority of methods discussed in Sections 6–8 as authors often omit such an estimation. This makes the overall comparison of methods in terms of computational complexity impossible.

Now let us discuss the hyperparameter tuning problem. It turns out that some authors tune hyperparameters in their methods manually, some just do not consider the problem at all. Another issue is the lack of a general understanding how to determine "equal impact" of structure and attributes on the community detection results. For example, within weight-based methods (Section 6.1) some authors choose $\alpha = 1/2$ in experiments hoping that this provides the equal impact. However, if ones takes into account the different nature of structural and attributive information and the disbalance between associated statistical features, this choice seems questionable in general.

Furthermore, authors use different datasets (of various size and nature, see Section 5) and quality measures to test their methods so that a unified comparison of experimental results in different papers cannot be carried out.[10] What is more, datasets and software implementations used in comparison experiments are rarely provided by the authors (especially of early papers) thus making reproducing their results time-consuming or even impossible. Note how few links to source codes are indicated in Sections 6–8.

Let us now discuss the methodology of using quality measures in comparison experiments. There exist two main strategies (see Section 5). Namely, community detection quality can be evaluated (a) by heuristic measures of structural closeness and attributes homogeneity (e.g. Modularity and Entropy) when the dataset under consideration has no "ground truth" and (b) by measures estimating the agreement between the detected communities and the ground truth ones (e.g. NMI or ARI).

Let us first discuss Option (a) in terms of a typical weight-based early fusion method from Section 6.1. For a given node-attributed graph G, suppose that we first convert its attributes $(\mathscr{V}, \mathscr{A})$ into attributive graph G_A and further fuse it with the structure $(\mathscr{V}, \mathscr{E})$. This results in weighted graph G_W that is thought to contain information about both the structure and the attributes in a unified form. After that, we find communities in G_W that provide, say, the maximum of Modularity of G_W (e.g. by Weighted Louvain). At the evaluation step we further calculate Modularity of $(\mathscr{V}, \mathscr{E})$ and Entropy of $(\mathscr{V}, \mathscr{A})$ basing on the partition found. This seems reasonable as far as you do not look at such a methodology critically. Indeed, note that we deal with one quality measure within the optimization process (Modularity of G_W) but evaluate community detection performance by other measures (Modularity of $(\mathscr{V}, \mathscr{E})$ and Entropy of $(\mathscr{V}, \mathscr{A})$). Generally speaking, how can we be sure that optimization of one objective function provides optimal values of other objective functions, if there is no mathematically established connection between them? Of course, this may be simply called a heuristic but it looks more a logical gap, from our point of view. Anyways, we are unaware of any explicit explanation of such a methodology. What is more, one should carefully study how the change of data representation within a method (the change of vectors $\mathscr{A}$ for edge weights in G_A in the above-mentioned example) affects community detection results.

Take into account that a similar discussion is suitable for the majority of methods in Sections 6–8 that use quality measures for estimating structural closeness and attributes homogeneity of the detected communities. The exception is the methods that directly optimize the quality measures *within the optimization process*. Examples are the simultaneous fusion methods **OCru** [49],

[10] A more general observation is that several comparison experiments clearly do not provide generality of conclusions. This however seems to be a hot topic among supporters of scientific research from one side and of empirical one from another side.

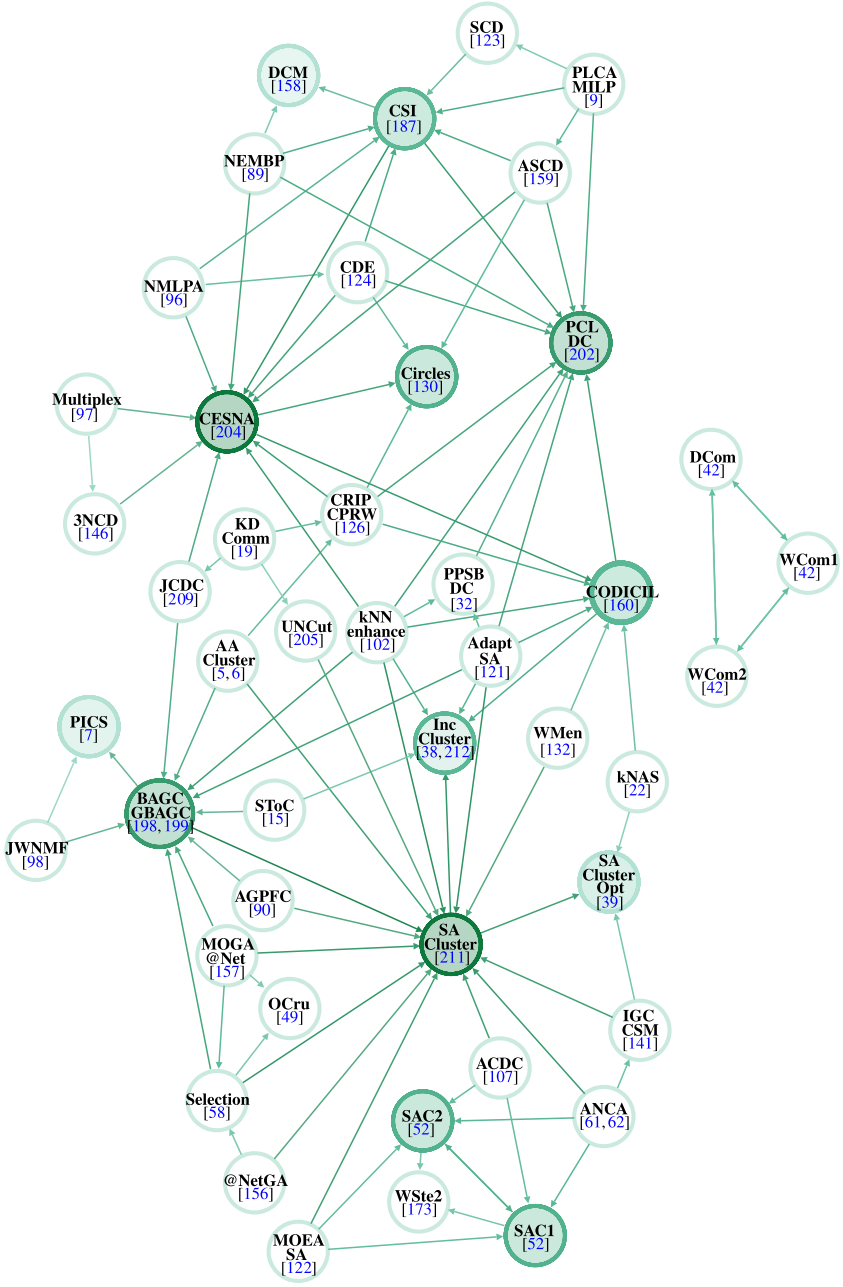

Fig. 8. The directed graph of existing method–method comparisons. Nodes (shown only those with degree $\geqslant 2$) represent the methods classified in the present survey. The most influential methods (nodes with highest PageRank) are filled green so that the darker green means the higher PageRank.

SAC1 [52], **ILouvain** [43], **UNCut** [205], **MOEA-SA** [122], **MOGA-@Net** [157] and **JCDC** [209]. Just by construction, they aim at providing optimal values of the quality measures. One should however take into account the precision of the optimization method applied.

Now we turn our attention to the methodology of evaluating community detection quality by measures that estimate the agreement between the detected communities and the ground truth ones (Option (b)). In our opinion, such a methodology makes sense for synthetic node-attributed social networks as the way how the communities are formed is known in advance. As for real-world networks, it seems somewhat questionable as ground truth communities may reflect only one point of view on network communities (among many possible). Therefore, expecting that any community detection method shares this point of view seems a bit unsuitable. There are several works on this issue, e.g. [92,144,152], discussing connections between ground truth, attributes and quality measures in detail, and therefore we refer an interested reader to them.

Recall that we started this discussion trying to determine the state-of-the-art methods for community detection in node-attributed social networks. Unfortunately, the above-mentioned facts do not give us a chance to do this. Of course, we could simply list the most recent methods in the field (and then it is enough to check just the date of publication), but this certainly does not meet the requirements imposed on state-of-the-art methods.

10. Conclusions

It is shown in the survey that there exist a large amount of methods for community detection in node-attribute social networks based on different ideas. In particular, we gave short

descriptions of 75 most relevant methods and mentioned much more of those partly related to the topic (Sections 6–8).

We also proposed to divide the methods into the three classes — early fusion, simultaneous fusion and late fusion ones (Section 4). This classification is based on the moment when network structure and attributes are fused within a method and allows a researcher/data scientist to estimate the ease of method's software implementation. Namely, we concluded that early and late fusion methods can be easier implemented in comparison with simultaneous fusion ones as the former two usually can be combined of several classical community detection algorithms with existing implementations. At a lower lever, we also divided the methods into subclasses of fusion techniques used (Section 4). It allows one to estimate the methodological variety in the field.

Within the classification, we also focused on the experiments performed so that one can see which datasets and measures are used for quality evaluation of each method from Sections 5–8.

The analysis of all the information collected brought us to the unfortunate conclusion that it is impossible now to determine state-of-the-art methods for community detection in node-attributed social networks (Section 9). This is a result of the presence of the following general problems in the field that we disclosed in the survey:

- the terminology in the field is rather unstable (Section 3.2);
- there is no generally accepted opinion on the effect of fusing structure and attributes, in particular, on when the fusion is helpful and when not in terms of subclasses of node-attributed social networks (Section 2.2);
- there is no unified methodology for experimental comparison of methods that would include estimation of computational complexity, use of a unified collection of datasets and quality measures for evaluation, and justified hyperparameter tuning procedures (Section 9);
- there is no general understanding what is the "equal impact" of structure and attributes on community detection results (Section 9);
- as a rule, there is no mathematically established connection between computational processes within a community detection method and the quality measures used for its evaluation (Section 9).

Summarizing, we concluded that the comparison study allowing one to determine the most preferable (in any sense) methods in the field is far from being complete.

Nevertheless, community detection methods dealing both with network structure and attributes remain a powerful tool in social network analysis and can yield useful insights to a researcher/data scientist. Furthermore, the methods have wide applications even beyond social networks (Section 3). With respect to these, we believe that the formulation of existing problems in the field done in this survey is the first step in finding solutions to them.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author is very grateful to Klavdiya Bochenina and Timofey Gradov for numerous conversations on the topic and especially to the Anonymous Referee for useful comments that helped to improve the quality of exposition in the survey essentially.

This research is financially supported by Russian Science Foundation, Agreement 17-71-30029 with co-financing of Bank Saint Petersburg, Russia.

References

[1] Lada A.A.damic, Eytan Adar, Friends and neighbors on the web, Social Networks 25 (3) (2003) 211–230.
[2] Lada A. Adamic, Natalie Glance, The political blogosphere and the 2004 U.S. election: Divided they blog, in: Proceedings of the 3rd International Workshop on Link Discovery, LinkKDD '05, ACM, New York, NY, USA, 2005, pp. 36–43.
[3] Charu C. Aggarwal, ChengXiang Zhai, A survey of text clustering algorithms, in: Mining Text Data, Springer US, Boston, MA, 2012, pp. 77–128.
[4] Yong-Yeol Ahn, James P. Bagrow, Sune Lehmann, Link communities reveal multiscale complexity in networks, Nature 466 (2010) 761–764.
[5] Esra Akbas, Peixiang Zhao, Attributed graph clustering: An attribute-aware graph embedding approach, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, ACM, New York, NY, USA, 2017, pp. 305–308.
[6] Esra Akbas, Peixiang Zhao, Graph clustering based on attribute-aware graph embedding, in: From Security To Community Detection in Social Networking Platforms, Springer International Publishing, Cham, 2019, pp. 109–131.
[7] Leman Akoglu, Hanghang Tong, Brendan Meeder, Christos Faloutsos, PICS: Parameter-free identification of cohesive subgroups in large attributed graphs, in: Proceedings of the 12th SIAM International Conference on Data Mining, SDM 2012, 2012, pp. 439–450.
[8] Andry Alamsyah, Budi Rahardjo, Kuspriyanto, Community detection methods in social network analysis, Adv. Sci. Lett. 20 (1) (2014) 250–253.
[9] Esmaeil Alinezhad, Babak Teimourpour, Mohammad Mehdi Sepehri, Mehrdad Kargari, Community detection in attributed networks considering both structural and attribute similarities: two mathematical programming approaches, Neural Comput. Appl. (2019).
[10] C. Ambroise, M. Dang, G. Govaert, Clustering of spatial data by the EM algorithm, in: Amílcar Soares, Jaime Gómez-Hernandez, Roland Froidevaux (Eds.), GeoENV I — Geostatistics for Environmental Applications, Springer Netherlands, Dordrecht, 1997, pp. 493–504.
[11] Yousra Asim, Rubina Ghazal, Wajeeha Naeem, Abdul Majeed, Basit Raza, Ahmad Kamran Malik, Community detection in networks using node attributes and modularity, Int. J. Adv. Comput. Sci. Appl. 8 (1) (2017).
[12] Martin Atzmueller, Stephan Doerfel, Folke Mitzlaff, Description-oriented community detection using exhaustive subgroup discovery, Inform. Sci. 329 (2016) 965–984, Special issue on Discovery Science.
[13] Martin Atzmueller, Henry Soldano, Guillaume Santini, Dominique Bouthinon, MinerLSD: efficient mining of local patterns on attributed networks, Appl. Netw. Sci. 4 (1) (2019) 43.
[14] Ramnath Balasubramanyan, William W. Cohen, Block-LDA: Jointly modeling entity-annotated text and entity-entity links, in: Proceedings of the 2011 SIAM International Conference on Data Mining, 2011, pp. 450–461.
[15] Alessandro Baroni, Alessio Conte, Maurizio Patrignani, Salvatore Ruggieri, Efficiently clustering very large attributed graphs, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, ACM, New York, NY, USA, 2017, pp. 369–376.
[16] Dominik Benz, Andreas Hotho, Robert Jäschke, Beate Krause, Folke Mitzlaff, Christoph Schmitz, Gerd Stumme, The social bookmark and publication management system bibsonomy, VLDB J. 19 (6) (2010) 849–875.
[17] M. Berlingerio, M. Coscia, F. Giannotti, Finding and characterizing communities in multidimensional networks, in: 2011 International Conference on Advances in Social Networks Analysis and Mining, 2011, pp. 490–494.
[18] Michele Berlingerio, Fabio Pinelli, Francesco Calabrese, ABACUS: frequent pattern mining-based community discovery in multidimensional networks, Data Min. Knowl. Discov. 27 (3) (2013) 294–320.
[19] Shreyansh Bhatt, Swati Padhee, Amit Sheth, Keke Chen, Valerie Shalin, Derek Doran, Brandon Minnery, Knowledge graph enhanced community detection and characterization, in: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM '19, ACM, New York, NY, USA, 2019, pp. 51–59.
[20] N. Binkiewicz, J.T. Vogelstein, K. Rohe, Covariate-assisted spectral clustering, Biometrika 104 (2) (2017) 361–377.
[21] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.
[22] M. Parimala Boobalan, Daphne Lopez, X.Z. Gao, Graph clustering using k-neighbourhood attribute structural similarity, Appl. Soft Comput. 47 (C) (2016) 216–223.
[23] Tossapon Boongoen, Natthakan Iam-On, Cluster ensembles: A survey of approaches with recent extensions and applications, Comp. Sci. Rev. 28 (2018) 1–25.
[24] Cecile Bothorel, Juan David Cruz, Matteo Magnani, Barbora Micenková, Clustering attributed graphs: Models, measures and methods, Netw. Sci. 3 (3) (2015) 408–444.

[25] Oualid Boutemine, Mohamed Bouguessa, Mining community structures in multidimensional networks, ACM Trans. Knowl. Discov. Data 11 (4) (2017) 51:1–51:36.

[26] Zhan Bu, Guangliang Gao, Hui-Jia Li, Jie Cao, CAMAS: A cluster-aware multiagent system for attributed graph clustering, Inf. Fusion 37 (2017) 10–21.

[27] Z. Bu, H. Li, J. Cao, Z. Wang, G. Gao, Dynamic cluster formation game for attributed graph clustering, IEEE Trans. Cybern. 49 (1) (2019) 328–341.

[28] Deng Cai, Xiaofei He, Xiaoyun Wu, Jiawei Han, Non-negative matrix factorization on manifold, in: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, ICDM '08, IEEE Computer Society, Washington, DC, USA, 2008, pp. 63–72.

[29] H. Cai, V.W. Zheng, K.C. Chang, A comprehensive survey of graph embedding: Problems, techniques, and applications, IEEE Trans. Knowl. Data Eng. 30 (9) (2018) 1616–1637.

[30] Xiangyong Cao, Xiangyu Chang, Zongben Xu, Community detection for clustered attributed graphs via a variational EM algorithm, in: Proceedings of the 2014 International Conference on Big Data Science and Computing, BigDataScience '14, ACM, New York, NY, USA, 2014, p. 28:1.

[31] Shaosheng Cao, Wei Lu, Qiongkai Xu, Grarep: Learning graph representations with global structural information, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15, ACM, New York, NY, USA, 2015, pp. 891–900.

[32] Bian-fang Chai, Jian Yu, Cai-yan Jia, Tian-bao Yang, Ya-wen Jiang, Combining a popularity-productivity stochastic block model with a discriminative-content model for general structure detection, Phys. Rev. E 88 (2013) 012807.

[33] Tanmoy Chakraborty, Ayushi Dalmia, Animesh Mukherjee, Niloy Ganguly, Metrics for community analysis: A survey, ACM Comput. Surv. 50 (4) (2017) 54:1–54:37.

[34] Tanmoy Chakraborty, Sriram Srinivasan, Niloy Ganguly, Animesh Mukherjee, Sanjukta Bhowmick, On the permanence of vertices in network communities, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, ACM, New York, NY, USA, 2014, pp. 1396–1405.

[35] Jonathan Chang, David Blei, Relational topic models for document networks, in: David van Dyk, Max Welling (Eds.), Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics, in: Proceedings of Machine Learning Research, vol. 5, PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 2009, pp. 81–88.

[36] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C. Aggarwal, Thomas S. Huang, Heterogeneous network embedding via deep architectures, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, ACM, New York, NY, USA, 2015, pp. 119–128.

[37] Chun-Hung Cheng, Ada Waichee Fu, Yi Zhang, Entropy-based subspace clustering for mining numerical data, in: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '99, ACM, New York, NY, USA, 1999, pp. 84–93.

[38] Hong Cheng, Yang Zhou, Xin Huang, Jeffrey Xu Yu, Clustering large attributed information networks: an efficient incremental computing approach, Data Min. Knowl. Discov. 25 (3) (2012) 450–477.

[39] Hong Cheng, Yang Zhou, Jeffrey Xu Yu, Clustering large attributed graphs: A balance between structural and attribute similarities, ACM Trans. Knowl. Discov. Data 5 (2) (2011) 12:1–12:33.

[40] Aaron Clauset, M.E.J. Newman, Cristopher Moore, Finding community structure in very large networks, Phys. Rev. E 70 (2004) 066111.

[41] David A. Cohn, Thomas Hofmann, The missing link - a probabilistic model of document content and hypertext connectivity, in: T.K. Leen, T.G. Dietterich, V. Tresp (Eds.), Advances in Neural Information Processing Systems, vol. 13, MIT Press, 2001, pp. 430–436.

[42] David Combe, Christine Largeron, Elod Egyed-Zsigmond, Mathias Gery, Combining relations and text in scientific network clustering, in: Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012, ASONAM '12, IEEE Computer Society, Washington, DC, USA, 2012, pp. 1248–1253.

[43] David Combe, Christine Largeron, Mathias Géry, Előd Egyed-Zsigmond, I-louvain: An attributed graph clustering method, in: Elisa Fromont, Tijl De Bie, Matthijs van Leeuwen (Eds.), Advances in Intelligent Data Analysis XIV, Springer International Publishing, Cham, 2015, pp. 181–192.

[44] Michele Coscia, Fosca Giannotti, Dino Pedreschi, A classification for community discovery methods in complex networks, Stat. Anal. Data Min. 4 (5) (2011) 512–546.

[45] Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom Mitchell, Kamal Nigam, Sean Slattery, Learning to extract symbolic knowledge from the world wide web, in: Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence, American Association for Artificial Intelligence, Menlo Park, 1998, pp. 509–516.

[46] R. Cross, A. Parker, The Hidden Power of Social Networks, Harvard Business School Press, Boston, MA, USA, 2004.

[47] J.D. Cruz, C. Bothorel, Information integration for detecting communities in attributed graphs, in: 2013 Fifth International Conference on Computational Aspects of Social Networks, 2013, pp. 62–67.

[48] Juan Cruz, Cécile Bothorel, François Poulet, Détection et visualisation des communautés dans les réseaux sociaux, Rev. Intell. Artif. 26 (2012) 369–392.

[49] Juan David Cruz Gomes, Cécile Bothorel, François Poulet, Semantic clustering of social networks using points of view, in: CORIA: Conférence en Recherche d'Information et Applications 2011, Avignon, France, 2011.

[50] Juan David Cruz Gomez, Cécile Bothorel, François Poulet, Entropy based community detection in augmented social networks, in: International Conference on Computational Aspects of Social Networks, Salamanca, Spain, 2011, pp. 163–168.

[51] P. Cui, X. Wang, J. Pei, W. Zhu, A survey on network embedding, IEEE Trans. Knowl. Data Eng. 31 (5) (2019) 833–852.

[52] The Anh Dang, Emmanuel Viennet, Community detection based on structural and attribute similarities, in: International Conference on Digital Society, ICDS, Jan. 2012, pp. 7–14, (ISBN: 978-1-61208-176-2). Best paper award.

[53] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1) (1977) 1–38.

[54] Karine Descormiers, Carlo Morselli, Alliances, conflicts, and contradictions in montreal's street gang landscape, Int. Crim. Justice Rev. 21 (3) (2011) 297–314.

[55] Inderjit S. Dhillon, Subramanyam Mallela, Dharmendra S. Modha, Information-theoretic co-clustering, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, ACM, New York, NY, USA, 2003, pp. 89–98.

[56] Ying Ding, Community detection: Topological vs. topical, J. Inform. 5 (4) (2011) 498–514.

[57] Nathan Eagle, Alex (Sandy) Pentland, David Lazer, Inferring friendship network structure by using mobile phone data, Proc. Natl. Acad. Sci. 106 (36) (2009) 15274–15278.

[58] Haithum Elhadi, Gady Agam, Structure and attributes community detection: Comparative analysis of composite, ensemble and selection methods, in: Proceedings of the 7th Workshop on Social Network Mining and Analysis, SNAKDD '13, ACM, New York, NY, USA, 2013, pp. 10:1–10:7.

[59] Elena Erosheva, Stephen Fienberg, John Lafferty, Mixed-membership models of scientific publications, Proc. Natl. Acad. Sci. 101 (Suppl. 1) (2004) 5220–5227.

[60] Martin Ester, Rong Ge, Byron J. Gao, Zengjian Hu, Boaz Ben-Moshe, Joint cluster analysis of attribute data and relationship data: the connected k-center problem, in: SDM, 2006.

[61] Issam Falih, Nistor Grozavu, Rushed Kanawati, Younes Bennani, Community detection in attributed network, in: WWW '18 Companion Proceedings of the the Web Conference 2018, 2018, pp. 1299–1306.

[62] Issam Falih, Nistor Grozavu, Rushed Kanawati, Younès Bennani, ANCA : Attributed network clustering algorithm, in: Chantal Cherifi, Hocine Cherifi, Márton Karsai, Mirco Musolesi (Eds.), Complex Networks & their Applications VI, Springer International Publishing, Cham, 2018, pp. 241–252.

[63] Saeed Farzi, Sahar Kianian, A novel clustering algorithm for attributed graphs based on k-medoid algorithm, J. Exp. Theor. Artif. Intell. 30 (6) (2018) 795–809.

[64] Andrew Fiore, Judith Donath, Homophily in online dating: When do you like someone like yourself? in: Conference on Human Factors in Computing Systems - Proceedings, 2005, pp. 1371–1374.

[65] Santo Fortunato, Community detection in graphs, Phys. Rep. 486 (3) (2010) 75–174.

[66] A.L.N. Fred, A.K. Jain, Data clustering using evidence accumulation, in: Object Recognition Supported By User Interaction for Service Robots, vol. 4, 2002, pp. 276–280.

[67] Brendan J. Frey, Delbert Dueck, Clustering by passing messages between data points, Science 315 (5814) (2007) 972–976.

[68] Hongchang Gao, Heng Huang, Deep attributed network embedding, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization, 2018, pp. 3364–3370.

[69] Jing Gao, Feng Liang, Wei Fan, Chi Wang, Yizhou Sun, Jiawei Han, On community outliers and their efficient detection in information networks, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10, ACM, New York, NY, USA, 2010, pp. 813–822.

[70] Rong Ge, Martin Ester, Byron J. Gao, Zengjian Hu, Binay Bhattacharya, Boaz Ben-Moshe, Joint cluster analysis of attribute data and relationship data: The connected k-center problem, algorithms and applications, ACM Trans. Knowl. Discov. Data 2 (2) (2008) 7:1–7:35.

[71] Lisa Getoor, Nir Friedman, Daphne Koller, Benjamin Taskar, Learning probabilistic models of link structure, J. Mach. Learn. Res. 3 (2003) 679–707.

[72] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. 99 (12) (2002) 7821–7826.

[73] Derek Greene, Pádraig Cunningham, Producing a unified graph representation from multiple social network views, in: Proceedings of the 5th Annual ACM Web Science Conference, WebSci '13, ACM, New York, NY, USA, 2013, pp. 118–121.

[74] Aditya Grover, Jure Leskovec, Node2vec: Scalable feature learning for networks, in: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, ACM, New York, NY, USA, 2016, pp. 855–864.

[75] Thomas U. Grund, James A. Densley, Ethnic homophily and triad closure: Mapping internal gang structure using exponential random graph models, J. Contemp. Crim. Justice 31 (3) (2015) 354–370.

[76] P.D. Grünwald, The Minimum Description Length Principle, The MIT Press, 2007.

[77] Quanquan Gu, Jie Zhou, Co-clustering on manifolds, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09, ACM, New York, NY, USA, 2009, pp. 359–368.

[78] Francesco Gullo, Carlotta Domeniconi, Andrea Tagarelli, Projective clustering ensembles, Data Min. Knowl. Discov. 26 (3) (2013) 452–511.

[79] Stephan Gunnemann, Subspace clustering for complex data, in: Volker Markl, Gunter Saake, Kai-Uwe Sattler, Gregor Hackenbroich, Bernhard Mitschang, Theo Harder, Veit Koppen (Eds.), Datenbanksysteme Fur Business, Technologie Und Web (BTW) 2034, Gesellschaft fur Informatik e.V., Bonn, 2013, pp. 343–362.

[80] Stephan Günnemann, Brigitte Boden, Ines Färber, Thomas Seidl, Efficient mining of combined subspace and subgraph clusters in graphs with feature vectors, in: Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, Guandong Xu (Eds.), Advances in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 261–275.

[81] Stephan Günnemann, Brigitte Boden, Thomas Seidl, DB-CSC: A density-based approach for subspace clustering in graphs with feature vectors, in: Proceedings of the 2011th European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part I, ECMLPKDD'11, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 565–580.

[82] S. Gunnemann, I. Farber, B. Boden, T. Seidl, Subspace clustering meets dense subgraph mining: A synthesis of two paradigms, in: 2010 IEEE International Conference on Data Mining, 2010, pp. 845–850.

[83] Stephan Günnemann, Ines Färber, Brigitte Boden, Thomas Seidl, GAMer: a synthesis of subspace clustering and dense subgraph mining, Knowl. Inf. Syst. 40 (2) (2014) 243–278.

[84] Stephan Günnemann, Ines Färber, Sebastian Raubach, Thomas Seidl, Spectral subspace clustering for graphs with feature vectors, in: 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 231–240.

[85] T. Guo, S. Pan, X. Zhu, C. Zhang, CFOND: Consensus factorization for co-clustering networked data, IEEE Trans. Knowl. Data Eng. 31 (4) (2019) 706–719.

[86] Will Hamilton, Zhitao Ying, Jure Leskovec, Inductive representation learning on large graphs, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017, pp. 1024–1034.

[87] J.A. Hartigan, M.A. Wong, A k-means clustering algorithm, J. R. Stat. Soc. Ser. C. Appl. Stat. 28 (1) (1979) 100–108.

[88] C. He, X. Fei, H. Li, Y. Tang, H. Liu, Q. Chen, A multi-view clustering method for community discovery integrating links and tags, in: 2017 IEEE 14th International Conference on E-Business Engineering, ICEBE, 2017, pp. 23–30.

[89] Dongxiao He, Zhiyong Feng, Di Jin, Xiaobao Wang, Weixiong Zhang, Joint identification of network communities and semantics via integrative modeling of network topologies and node contents, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI'17, AAAI Press, 2017, pp. 116–124.

[90] Chaobo He, Shuangyin Liu, Lei Zhang, Jianhua Zheng, A fuzzy clustering based method for attributed graph partitioning, J. Ambient Intell. Humanized Comput. 10 (9) (2019) 3399–3407.

[91] Paul W. Holland, Kathryn Blackmond Laskey, Samuel Leinhardt, Stochastic blockmodels: First steps, Social Networks 5 (2) (1983) 109–137.

[92] Darko Hric, Richard K. Darst, Santo Fortunato, Community detection in networks: Structural communities versus ground truth, Phys. Rev. E 90 (2014) 062805.

[93] L. Hu, K.C.C. Chan, Fuzzy clustering in a complex network based on content relevance and link structures, IEEE Trans. Fuzzy Syst. 24 (2) (2016) 456–470.

[94] Xiao Huang, Jundong Li, Xia Hu, Label informed attributed network embedding, in: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM '17, ACM, New York, NY, USA, 2017, pp. 731–739.

[95] Zhipeng Huang, Nikos Mamoulis, Heterogeneous information network embedding for meta path based proximity, 2017, arXiv arXiv:abs/1701.05291.

[96] Bingyang Huang, Chaokun Wang, Binbin Wang, NMLPA: Uncovering overlapping communities in attributed networks via a multi-label propagation approach, Sensors (Basel, Switzerland) 19 (2) (2019) 260.

[97] Y. Huang, H. Wangg, Consensus and multiplex approach for community detection in attributed networks, in: 2016 IEEE Global Conference on Signal and Information Processing, GlobalSIP, 2016, pp. 425–429.

[98] Zhichao Huang, Yunming Ye, Xutao Li, Feng Liu, Huajie Chen, Joint weighted nonnegative matrix factorization for mining attributed graphs, in: Jinho Kim, Kyuseok Shim, Longbing Cao, Jae-Gil Lee, Xuemin Lin, Yang-Sae Moon (Eds.), Advances in Knowledge Discovery and Data Mining, Springer International Publishing, Cham, 2017, pp. 368–380.

[99] Roberto Interdonato, Martin Atzmueller, Sabrina Gaito, Rushed Kanawati, Christine Largeron, Alessandra Sala, Feature-rich networks: going beyond complex network topologies, Appl. Netw. Sci. 4 (1) (2019) 4.

[100] Hiroyoshi Ito, Takahiro Komamizu, Toshiyuki Amagasa, Hiroyuki Kitagawa, Community detection and correlated attribute cluster analysis on multi-attributed graphs, in: EDBT/ICDT Workshops, 2018.

[101] Tomoharu Iwata, Kazumi Saito, Naonori Ueda, Sean Stromsten, Thomas L. Griffiths, Joshua B. Tenenbaum, Parametric embedding for class visualization, Neural Comput. 19 (9) (2007) 2536–2556.

[102] Caiyan Jia, Yafang Li, Matthew B. Carson, Xiaoyang Wang, Jian Yu, Node attribute-enhanced community detection in complex networks, Sci. Rep. 7:2626 (2017) 1–15.

[103] Jianbo Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.

[104] Stephen C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (3) (1967) 241–254.

[105] D.R. Karger, Global min-cuts in RNC, and other ramifications of a simple min-cut algorithm, in: Proc. 4th Annual ACM-SIAM Symposium on Discrete Algorithms, 1993, pp. 21–30.

[106] George Karypis, Vipin Kumar, A fast and high quality multilevel scheme for partitioning irregular graphs, SIAM J. Sci. Comput. 20 (1) (1998) 359–392.

[107] N. Khediri, W. Karoui, Community detection in social network with node attributes based on formal concept analysis, in: 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications, AICCSA, Oct. 2017, pp. 1346–1353.

[108] Thomas N. Kipf, Max Welling, Variational graph auto-encoders, 2016, arXiv arXiv:abs/1611.07308.

[109] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, Mason A. Porter, Multilayer networks, J. Complex Netw. 2 (3) (2014) 203–271.

[110] Gueorgi Kossinets, Duncan J. Watts, Origins of homophily in an evolving social network, Am. J. Sociol. 115 (2009) 405–450.

[111] Andrea Lancichinetti, Santo Fortunato, Consensus clustering in complex networks, Sci. Rep. 2:336 (2012) 1–7.

[112] Andrea Lancichinetti, Santo Fortunato, János Kertész, Detecting the overlapping and hierarchical community structure in complex networks, New J. Phys. 11 (3) (2009) 033015.

[113] Andrea Lancichinetti, Santo Fortunato, Filippo Radicchi, Benchmark graphs for testing community detection algorithms, Phys. Rev. E 78 (2008) 046110.

[114] E. Lazega, The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership, Oxford University Press, Oxford, UK, 2001.

[115] T. M. V. Le, H. W. Lauw, Probabilistic latent document network embedding, in: 2014 IEEE International Conference on Data Mining, 2014, pp. 270–279.

[116] Daniel D. Lee, H. Sebastian Seung, Algorithms for non-negative matrix factorization, in: T.K. Leen, T.G. Dietterich, V. Tresp (Eds.), Advances in Neural Information Processing Systems, vol. 13, MIT Press, 2001, pp. 556–562.

[117] Jure Leskovec, Kevin J. Lang, Michael Mahoney, Empirical comparison of algorithms for network community detection, in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, 2010, pp. 631–640.

[118] Jure Leskovec, Julian J. Mcauley, Learning to discover social circles in ego networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, vol. 25, Curran Associates, Inc., 2012, pp. 539–547.

[119] Jundong Li, Ruocheng Guo, Chenghao Liu, Huan Liu, Adaptive unsupervised feature selection on attributed networks, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 92–100.

[120] P. Li, L. Huang, C. Wang, D. Huang, J. Lai, Community detection using attribute homogenous motif, IEEE Access 6 (2018) 47707–47716.

[121] Y. Li, C. Jia, X. Kong, L. Yang, J. Yu, Locally weighted fusion of structural and attribute information in graph clustering, IEEE Trans. Cybern. 49 (1) (2019) 247–260.

[122] Z. Li, J. Liu, K. Wu, A multiobjective evolutionary algorithm based on structural and attribute similarities for community detection in attributed networks, IEEE Trans. Cybern. 48 (7) (2018) 1963–1976.

[123] Zhen Li, Zhisong Pan, Guyu Hu, Guopeng Li, Xingyu Zhou, Detecting semantic communities in social networks, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. E100.A (11) (2017) 2507–2512.

[124] Ye Li, Chaofeng Sha, Xin Huang, Yanchun Zhang, Community detection in attributed graphs: An embedding approach, in: AAAI, 2018.

[125] Wu-Jun Li, Dit-Yan Yeung, Zhihua Zhang, Generalized latent factor models for social network analysis, in: IJCAI, 2011.

[126] L. Liu, L. Xu, Z. Wangy, E. Chen, Community detection based on structure and content: A content propagation perspective, in: 2015 IEEE International Conference on Data Mining, Nov. 2015, pp. 271–280.

[127] Sheng Luo, Zhifei Zhang, Yuanjian Zhang, Shuwen Ma, Co-association matrix-based multi-layer fusion for community detection in attributed networks, Entropy 21 (1) (2019).

[128] Gregory R. Madey, Albert-László Barabási, Nitesh V. Chawla, Marta Gonzalez, David Hachen, Brett Lantz, Alec Pawling, Timothy Schoenharl, Gábor Szabó, Pu Wang, Ping Yan, Enhanced situational awareness: Application of DDDAS concepts to emergency and disaster management, in: Yong Shi, Geert Dick van Albada, Jack Dongarra, Peter M.A. Sloot (Eds.), Computational Science – ICCS 2007, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 1090–1097.

[129] Peter V. Marsden, Noah E. Friedkin, Network studies of social influence, Sociol. Methods Res. 22 (1) (1993) 127–151.

[130] Julian Mcauley, Jure Leskovec, Discovering social circles in ego networks, ACM Trans. Knowl. Discov. Data 8 (1) (2014) 4:1–4:28.

[131] Miller McPherson, Lynn Smith-Lovin, James M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. 27 (1) (2001) 415–444.

[132] Fanrong Meng, Xiaobin Rui, Zhixiao Wang, Yan Xing, Longbing Cao, Coupled node similarity learning for community detection in attributed networks, Entropy 20 (6) (2018).

[133] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: Simple building blocks of complex networks, Science 298 (5594) (2002) 824–827.

[134] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee, Measurement and analysis of online social networks, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC '07, ACM, New York, NY, USA, 2007, pp. 29–42.

[135] Flavia Moser, Recep Colak, Arash Rafiey, Martin Ester, Mining cohesive patterns from graphs with feature vectors, in: SDM, SIAM, 2009, pp. 593–604.

[136] Flavia Moser, Rong Ge, Martin Ester, Joint cluster analysis of attribute and relationship data withouta-priori specification of the number of clusters, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '07, ACM, New York, NY, USA, 2007, pp. 510–519.

[137] E. Muller, P.I. Sanchez, Y. Mulle, K. Bohm, Ranking outlier nodes in subspaces of attributed graphs, in: 2013 IEEE 29th International Conference on Data Engineering Workshops, ICDEW 2013, IEEE Computer Society, Los Alamitos, CA, USA, 2013, pp. 216–222.

[138] N. Muslim, A combination approach to community detection in social networks by utilizing structural and attribute data, Soc. Network. 5 (2016) 11–15.

[139] M. P. Naik, H. B. Prajapati, V. K. Dabhi, A survey on semantic document clustering, in: 2015 IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT, 2015, pp. 1–10.

[140] Ramesh M. Nallapati, Amr Ahmed, Eric P. Xing, William W. Cohen, Joint latent topic models for text and citations, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08, ACM, New York, NY, USA, 2008, pp. 542–550.

[141] Waqas Nawaz, Kifayat-Ullah Khan, Young-Koo Lee, Sungyoung Lee, Intra graph clustering using collaborative similarity measure, Distrib. Parallel Databases 33 (4) (2015) 583–603.

[142] Jennifer Neville, Micah Adler, David Jensen, Clustering relational data using attribute and link information, in: Proceedings of the Text Mining and Link Analysis Workshop, 18th International Joint Conference on Artificial Intelligence, 2003, pp. 9–15.

[143] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (2006) 036104.

[144] M. Newman, Aaron Clauset, Structure and inference in annotated networks, Nature Commun. 7 (2015).

[145] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113.

[146] H.T. Nguyen, T.N. Dinh, Unveiling the structure of multi-attributed networks via joint non-negative matrix factorization, in: MILCOM 2015 - 2015 IEEE Military Communications Conference, Oct. 2015, pp. 1379–1384.

[147] Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj, Exploratory Social Network Analysis with Pajek, Cambridge University Press, New York, NY, USA, 2004.

[148] Madalina Olteanu, Nathalie Villa-Vialaneix, Christine Cierco-Ayrolles, Multiple kernel self-organizing maps, in: Verleysen, M. (Ed.), European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 2013, p. 83.

[149] Andreas Papadopoulos, George Pallis, Marios D. Dikaiakos, Weighted clustering of attributed multi-graphs, Computing 99 (9) (2017) 813–840.

[150] Andreas Papadopoulos, Dimitrios Rafailidis, George Pallis, Marios D. Dikaiakos, Clustering attributed multi-graphs with information ranking, in: Qiming Chen, Abdelkader Hameurlain, Farouk Toumani, Roland Wagner, Hendrik Decker (Eds.), Database and Expert Systems Applications, Springer International Publishing, Cham, 2015, pp. 432–446.

[151] M. Parimala, Daphne Lopez, Graph clustering based on Structural Attribute Neighborhood Similarity (SANS), in: 2015 IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT, 2015, pp. 1–4.

[152] Leto Peel, Daniel B. Larremore, Aaron Clauset, The ground truth about metadata and community detection in networks, Sci. Adv. 3 (5) (2017).

[153] Yulong Pei, Nilanjan Chakraborty, Katia Sycara, Nonnegative matrix tri-factorization with graph regularization for community detection in social networks, in: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, AAAI Press, 2015, pp. 2083–2089.

[154] Z. Pei, X. Zhang, F. Zhang, B. Fang, Attributed multi-layer network embedding, in: 2018 IEEE International Conference on Big Data, Big Data, Dec. 2018, pp. 3701–3710.

[155] Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sánchez, Emmanuel Müller, Focused clustering and outlier detection in large attributed graphs, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, ACM, New York, NY, USA, 2014, pp. 1346–1355.

[156] Clara Pizzuti, Annalisa Socievole, A genetic algorithm for community detection in attributed graphs, in: Kevin Sim, Paul Kaufmann (Eds.), Applications of Evolutionary Computation, Springer International Publishing, Cham, 2018, pp. 159–170.

[157] C. Pizzuti, A. Socievole, Multiobjective optimization and local merge for clustering attributed graphs, IEEE Trans. Cybern. (2019) 1–13.

[158] Simon Pool, Francesco Bonchi, Matthijs van Leeuwen, Description-driven community detection, ACM Trans. Intell. Syst. Technol. 5 (2) (2014) 28:1–28:28.

[159] Meng Qin, Di Jin, Kai Lei, Bogdan Gabrys, Katarzyna Musial-Gabrys, Adaptive community detection incorporating topology and content in social networks, Knowl.-Based Syst. 161 (2018) 342–356.

[160] Yiye Ruan, David Fuhry, Srinivasan Parthasarathy, Efficient community detection in large networks using content and links, in: Proceedings of the 22nd International Conference on World Wide Web, ACM, New York, NY, USA, 2013, pp. 1089–1098,

[161] Mrinmaya Sachan, Danish Contractor, Tanveer A. Faruquie, L. Venkata Subramaniam, Using content and interactions for discovering communities in social networks, in: Proceedings of the 21st International Conference on World Wide Web, WWW '12, ACM, New York, NY, USA, 2012, pp. 331–340.

[162] N. Y. Saiyad, H. B. Prajapati, V. K. Dabhi, A survey of document clustering using semantic approach, in: 2016 International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT, 2016, pp. 2555–2562.

[163] Patricia Iglesias Sánchez, Emmanuel Müller, Uwe Leo Korn, Klemens Böhm, Andrea Kappes, Tanja Hartmann, Dorothea Wagner, Efficient algorithms for a robust modularity-driven clustering of attributed graphs, in: SDM, 2015.

[164] P. I. Sanchez, E. Muller, F. Laforet, F. Keller, K. Bohm, Statistical selection of congruent subspaces for mining attributed graphs, in: 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 647–656.

[165] Venu Satuluri, Srinivasan Parthasarathy, Scalable graph clustering using stochastic flows: Applications to community discovery, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, 2009, pp. 737–746,

[166] Satu Elisa Schaeffer, Graph clustering, Comp. Sci. Rev. 1 (1) (2007) 27–64.

[167] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, Tina Eliassi-Rad, Collective classifiction in network data, AI Mag. 29 (2008) 93–106.

[168] Nasrullah Sheikh, Zekarias Kefato, Alberto Montresor, Gat2vec: representation learning for attributed graphs, Computing 101 (3) (2019) 187–209.

[169] Motoki Shiga, Ichigaku Takigawa, Hiroshi Mamitsuka, A spectral clustering approach to optimally combining numericalvectors with a modular network, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '07, ACM, New York, NY, USA, 2007, pp. 647–656.

[170] Tom A.B. Snijders, Krzysztof Nowicki, Estimation and prediction for stochastic blockmodels for graphs with latent block structure, J. Classification 14 (1997) 75–100.

[171] Benno Stein, Oliver Niggemann, On the nature of structure and its identification, in: Proceedings of the 25th International Workshop on Graph-Theoretic Concepts in Computer Science, in: WG '99, Springer-Verlag, London, UK, UK, 1999, pp. 122–134.

[172] Karsten Steinhaeuser, Nitesh V. Chawla, Community detection in a large real-world social network, in: Huan Liu, John J. Salerno, Michael J. Young (Eds.), Social Computing, Behavioral Modeling, and Prediction, Springer US, Boston, MA, 2008, pp. 168–175.

[173] Karsten Steinhaeuser, Nitesh V. Chawla, Identifying and evaluating community structure in complex networks, Pattern Recognit. Lett. 31 (5) (2010) 413–421.

[174] Alexander Strehl, Joydeep Ghosh, Cluster ensembles — a knowledge reuse framework for combining multiple partitions, J. Mach. Learn. Res. 3 (2003) 583–617.

[175] Y. Sun, J. Han, J. Gao, Y. Yu, iTopicModel: Information network-integrated topic modeling, in: 2009 Ninth IEEE International Conference on Data Mining, 2009, pp. 493–502.

[176] Andrea Tagarelli, Alessia Amelio, Francesco Gullo, Ensemble-based community detection in multilayer networks, Data Min. Knowl. Discov. 31 (5) (2017) 1506–1543.

[177] Aditya Tandon, Aiiad Albeshri, Vijey Thayananthan, Wadee Alhalabi, Santo Fortunato, Fast consensus clustering in complex networks, Phys. Rev. E 99 (2019) 042301.

[178] Lei Tang, Huan Liu, Scalable learning of collective behavior based on sparse social dimensions, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09, ACM, New York, NY, USA, 2009, pp. 1107–1116.

[179] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, Qiaozhu Mei, LINE: Large-scale information network embedding, in: Proceedings of the 24th International Conference on World Wide Web, WWW '15, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2015, pp. 1067–1077.

[180] Mariano Tepper, Guillermo Sapiro, From local to global communities in large networks through consensus, in: Alvaro Pardo, Josef Kittler (Eds.), Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Springer International Publishing, Cham, 2015, pp. 659–666.

[181] Fei Tian, Bin Gao, Qing Cui, Enhong Chen, Tie-Yan Liu, Learning deep representations for graph clustering, in: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, AAAI'14, AAAI Press, 2014, pp. 1293–1299.

[182] Yuanyuan Tian, Richard A. Hankins, Jignesh M. Patel, Efficient aggregation for graph summarization, in: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, SIGMOD '08, ACM, New York, NY, USA, 2008, pp. 567–580.

[183] H. Tong, C. Faloutsos, J. Pan, Fast random walk with restart and its applications, in: Sixth International Conference on Data Mining, ICDM'06, 2006, pp. 613–622.

[184] A.L. Traud, E.D. Kelsic, P.J. Mucha, M.A. Porter, Comparing community structure to characteristics in online collegiate social networks, SIAM Rev. 53 (3) (2011) 526–543.

[185] Amanda L. Traud, Peter J. Mucha, Mason A. Porter, Social structure of facebook networks, Physica A 391 (16) (2012) 4165–4180.

[186] Nathalie Villa-Vialaneix, Madalina Olteanu, Christine Cierco-Ayrolles, Carte auto-organisatrice pour graphes étiquetés, in: Atelier Fouilles de Grands Graphes (FGG) - EGC'2013, Toulouse, France, 2013, p. 4.

[187] Xiao Wang, Di Jin, Xiaochun Cao, Liang Yang, Weixiong Zhang, Semantic community identification in large attribute networks, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, AAAI Press, 2016, pp. 265–271.

[188] Hua Wang, Feiping Nie, Heng Huang, Fillia Makedon, Fast nonnegative matrix tri-factorization for large-scale data co-clustering, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two, IJCAI'11, AAAI Press, 2011, pp. 1553–1558.

[189] Chun Wang, Shirui Pan, Guodong Long, Xingquan Zhu, Jing Jiang, MGAE: Marginalized graph autoencoder for graph clustering, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17, ACM, New York, NY, USA, 2017, pp. 889–898.

[190] X. Wang, L. Tang, H. Gao, H. Liu, Discovering overlapping groups in social media, in: 2010 IEEE International Conference on Data Mining, 2010, pp. 569–578.

[191] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, 1994.

[192] J.J. Whang, D.F. Gleich, I.S. Dhillon, Overlapping community detection using neighborhood-inflated seed expansion, IEEE Trans. Knowl. Data Eng. 28 (5) (2016) 1272–1284.

[193] Peng Wu, Li Pan, Mining application-aware community organization with expanded feature subspaces from concerned attributes in social networks, Knowl.-Based Syst. 139 (2018) 1–12.

[194] Rongkai Xia, Yan Pan, Lei Du, Jian Yin, Robust multi-view spectral clustering via low-rank and sparse decomposition, in: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, AAAI'14, AAAI Press, 2014, pp. 2149–2155.

[195] Jierui Xie, Boleslaw K. Szymanski, Towards linear time overlapping community detection in social networks, in: Pang-Ning Tan, Sanjay Chawla, Chin Kuan Ho, James Bailey (Eds.), Advances in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 25–36.

[196] Zhiqiang Xu, James Cheng, Xiaokui Xiao, Ryohei Fujimaki, Yusuke Muraoka, Efficient nonparametric and asymptotic Bayesian model selection methods for attributed graph clustering, Knowl. Inf. Syst. 53 (1) (2017) 239–268.

[197] Zhiqiang Xu, Yiping Ke, Effective and efficient spectral clustering on text and link data, in: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, ACM, New York, NY, USA, 2016, pp. 357–366.

[198] Z. Xu, Y. Ke, Y. Wang, H. Cheng, J. Cheng, A model-based approach to attributed graph clustering, in: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, 2012, pp. 505–516.

[199] Zhiqiang Xu, Yiping Ke, Yi Wang, Hong Cheng, James Cheng, GBAGC: A general Bayesian framework for attributed graph clustering, ACM Trans. Knowl. Discov. Data 9 (1) (2014) 5:1–5:43.

[200] Zhao Yang, René Algesheimer, Claudio J. Tessone, A comparative analysis of community detection algorithms on artificial networks, Sci. Rep. (2016).

[201] Tianbao Yang, Yun Chi, Shenghuo Zhu, Yihong Gong, Rong Jin, Directed network community detection: A popularity and productivity link model, in: Proceedings of the 2010 SIAM International Conference on Data Mining, 2010, pp. 742–753.

[202] Tianbao Yang, Rong Jin, Yun Chi, Shenghuo Zhu, Combining link and content for community detection: A discriminative approach, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09, ACM, New York, NY, USA, 2009, pp. 927–936.

[203] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, Edward Y. Chang, Network representation learning with rich text information, in: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, AAAI Press, 2015, pp. 2111–2117.

[204] Jaewon Yang, Julian J. McAuley, Jure Leskovec, Community detection in networks with node attributes, in: 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 1151–1156.

[205] Wei Ye, Linfei Zhou, Xin Sun, Claudia Plant, Christian Böhm, Attributed graph clustering with unimodal normalized cut, in: Michelangelo Ceci, Jaakko Hollmén, Ljupčo Todorovski, Celine Vens, Sašo Džeroski (Eds.), Machine Learning and Knowledge Discovery in Databases, Springer International Publishing, Cham, 2017, pp. 601–616.

[206] T. Yoshida, Toward finding hidden communities based on user profile, in: 2010 IEEE International Conference on Data Mining Workshops, Dec. 2010, pp. 380–387.

[207] Donghua Yu, Guojun Liu, Maozu Guo, Xiaoyan Liu, An improved k-medoids algorithm based on step increasing and optimizing medoids, Expert Syst. Appl. 92 (2018) 464–473.

[208] Hugo Zanghi, Stevenn Volant, Christophe Ambroise, Clustering based on random graph model embedding vertex features, Pattern Recognit. Lett. 31 (9) (2010) 830–836.

[209] Yuan Zhang, Elizaveta Levina, Ji Zhu, Community detection in networks with node features, Electron. J. Statist. 10 (2) (2016) 3153–3178.

[210] Tong Zhang, Alexandrin Popescul, Byron Dom, Linear prediction models with graph regularization for web-page categorization, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06, ACM, New York, NY, USA, 2006, pp. 821–826.

[211] Yang Zhou, Hong Cheng, Jeffrey Xu Yu, Graph clustering based on structural/attribute similarities, Proc. VLDB Endow. 2 (1) (2009) 718–729.

[212] Yang Zhou, Hong Cheng, Jeffrey Xu Yu, Clustering large attributed graphs: An efficient incremental approach, in: Proceedings of the 2010 IEEE International Conference on Data Mining, ICDM '10, IEEE Computer Society, Washington, DC, USA, 2010, pp. 689–698.

[213] Ding Zhou, Eren Manavoglu, Jia Li, C. Lee Giles, Hongyuan Zha, Probabilistic models for discovering e-communities, in: Proceedings of the 15th International Conference on World Wide Web, WWW '06, ACM, New York, NY, USA, 2006, pp. 173–182.

[214] Shenghuo Zhu, Kai Yu, Yun Chi, Yihong Gong, Combining content and link for classification using matrix factorization, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '07, ACM, New York, NY, USA, 2007, pp. 487–494.