METHODOLOGIES AND APPLICATION

CrossMark

A novel granular approach for detecting dynamic online communities in social network

Hamideh Sadat Cheraghchi[1] · Ali Zakerolhosseini[1] · Saeed Bagheri Shouraki[2] · Elaheh Homayounvala[3]

Abstract

The great surge in the research of community discovery in complex network is going on due to its challenging aspects. Dynamicity and overlapping nature are among the common characteristics of these networks which are the main focus of this paper. In this research, we attempt to approximate the granular human-inspired viewpoints of the networks. This is especially helpful when making decisions with partial knowledge. In line with the principle of granular computing, in which precision is avoided, we define the micro- and macrogranules in two levels of nodes and communities, respectively. The proposed algorithm takes microgranules as input and outputs meaningful communities in rough macrocommunity form. For this purpose, the microgranules are drawn toward each other based on a new rough similarity measure defined in this paper. As a result, the structure of communities is revealed and adapted over time, according to the interactions observed in the network, and the number of communities is extracted automatically. The proposed model can deal with both the low and the sharp changes in the network. The algorithm is evaluated in multiple dynamic datasets and the results confirm the superiority of the proposed algorithm in various measures and scenarios.

Keywords Social network analysis · Dynamic community detection · Granular clustering · Evolutionary clustering

1 Introduction and motivation

The widespread usage of various internet-based platforms has provided new forms of social interaction and collaboration, which contributes to the creation of virtual communities. With rapid growth of online dynamic social networks, where users' joining in and withdrawing from communities is common, the study of recognizing dense group of entities is a very valuable research topic and has triggered the interest of a variety of groups from social scientists and economists, to physicians and politicians. Recognizing niche markets for targeted advertisements and detection of influential individuals in politics are among a few applications of the "dynamic community detection" research domain, which is studied in this paper.

In spite of the increased number of studies in the past decade on dynamic communities (Cazabet and Amblard 2014), the field still requires the special attention of researchers due to its challenging aspects. Today, people join multiple groups introduced by the circle of their friends and may stay for a while or leave the groups soon thereafter. There should be smooth changes in the discovered communities, where many individuals stay for a long time while accounting for unforeseen dramatic shifts in these communities due to external reasons. The desired approach to handle this task should be computationally feasible and rapid, in order to adapt to changes. Incremental dynamic community detection exploits the information of past time steps to estimate the current community structure and helps to produce smooth results. Evolutionary clustering is one dominant incremental paradigm widely used which accounts for the historical data of the network. The approach use the information of past time steps to improve the clustering quality and also minimize the community drift compared to previous time steps to produce smooth results. A recent survey (Hartmann et al. 2016) studied the application of this paradigm with differ-

✉ Hamideh Sadat Cheraghchi
   h.s.cheraghchi@gmail.com

1   Department of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran

2   Electrical Engineering Department, Sharif University of Technology, Tehran, Iran

3   Cyberspace Research Institute, Shahid Beheshti University, Tehran, Iran

ent approaches. However, optimization of these two criteria together makes the evolutionary clustering suitable for the network with small changes and prevents the application of this method to abrupt changes. On the other hand, most of these methods are designed for non-overlapping networks (Hartmann et al. 2016). The requirement to specify the number of communities is another problem observed in most studies (Chi et al. 2007; Tang et al. 2012). Finally, all the methods requiring high amount of computation or space are precluded in real dynamic community detection scenarios (Xu et al. 2014).

The goal of our research is to present a novel incremental approach for dynamic networks which can produce high-quality communities while being able to handle the abrupt changes in the network. On the other hand, the approach should estimate the number of communities automatically and functions locally. Finally, dealing with uncertainty in categorization is another important aspect considered, which is also considered in this study. For this, we take inspiration from the following human behaviors:

1. Human beings have a granular view of the world, thus prohibiting precise boundaries;
2. Human decision-making is mostly intertwined with uncertainty and biased toward the circle of one's friends; and
3. They are able to categorize both the elements similar to the previous patterns stored in their minds and also allowing them to integrate with the new patterns by their representative-based categorization model (Grossberg 2013).

Using these principles, we propose our novel community detection algorithm called *Granular-ARTISON* to achieve the above-mentioned goals. Shortly, the contribution of the proposed algorithm can be listed as follows:

- It introduces a novel social network modeling based on granular concepts.
- It presents a new procedure for dynamic incremental clustering algorithms capable of detecting both low and abrupt changes in the network.
- It provides an explicit novel formula for embracing uncertainty in similarity matching process.
- It functions locally and is applicable to both weighted and binary networks.
- It outperforms the existing dynamic state-of-the-art evolutionary clustering algorithms in several experiments and various internal/external measures.

To elaborate the functionality and the performance of the proposed algorithm, the paper is outlined as follows. In Sect. 2, we review several related methods and background knowledge in two subsections of crisp dynamic and soft community detection algorithms. Section 3 explains our proposed approach and Sect. 4 represents the experiments to evaluate the proposed algorithm. Finally, we present our conclusions and future directions in Sect. 5.

2 Related work

2.1 Crisp community detection in temporal networks

The applicability of community detection in different domains has stimulated this research domain during the past decades. According to research papers on the static community detection domain (Plantié and Crampes 2013), the main categories are enumerated as follows: (1) partitioning methods, (2) hierarchical methods, (3) modularity optimization methods, (4) inference-based algorithms, (5) spectral methods, and (6) label propagation methods.

Mainly, community detection algorithms in temporal networks follow two main lines of research (Cazabet and Amblard 2014). In one prominent category, called *independent* community detection approach, the community mining process is performed in each snapshot of the network separately. Then, some correspondence with communities is obtained using similarity/distance measures to determine the relationship among them (Rosvall and Bergstrom 2010). The best advantage offered in this approach is the possibility of reusing traditional clustering methods in each snapshot. However, since most algorithms are seed-based, the emergence of unstable communities that differ drastically in different snapshots is the main weakness of this approach.

In the other category, called *incremental community detection,* the algorithms incorporate the information obtained in other snapshots for extracting communities of the current time step. This approach improves the time and computational complexity compared to the independent community detection approach (Takaffoli et al. 2011). An important series of works proposed in this category are *evolutionary* community detection algorithms, in which two potentially contradicting criteria or cost functions in an additive equation should be optimized (Chakrabarti et al. 2006). The first term relates to the correspondence of clustering results to current data as much as possible (*clustering quality*), and the second involves keeping the shifts of the results between current clustering and the previous time step as low as possible (*history quality*) to allow for *temporal smoothness* between clustering results in consecutive time steps. A weighting parameter α $(0 < \alpha < 1)$ (also known as smoothing factor) is used to weigh the two potentially contradicting criteria (Eq. (1)). Hence, the approach is best matched with networks with low changes, because there is deteriorating clustering quality in networks with rapid changes due to additive cost function.

$$\text{Cost}_{\text{total}} = \alpha SC(G_t, C_t) + (1 - \alpha) TC(C_{t-1}, C_t) \qquad (1)$$

where $SC(G_t, C_t)$ (SC for *snapshot cost*) measures the quality of the community C_t and $TC(C_{t-1}, C_t)$ (TC for *temporal cost*) measures the drift of the community discovered in two consecutive time steps. Several improved frameworks like *AFFECT* (Xu et al. 2014) are proposed, where the weighting factor is estimated using a statistical method according to the observed changes in the network.

Amendment of traditional static clustering models using the incremental approach is introduced in the literature in different categories of partitioning (Chakrabarti et al. 2006), agglomerative hierarchical clustering such as modularity-based clustering (Görke et al. 2013), spectral-based clustering (Xu et al. 2014), label propagation-based algorithms (Xie et al. 2013), and inference-based algorithms (Lin et al. 2009).

2.2 Soft community detection

One can recognize two broad categories of node-based (Xie et al. 2013) and link-based algorithms that address the problem of multiple belongingness of nodes to different communities in social networks (Ding et al. 2016). In a more detailed view (Xie et al. 2013), node-based methods are subcategorized into seed-based and local expansion algorithms (Whang et al. 2016), clique expansion algorithms like CPM (Palla et al. 2005) as the pioneer of overlapping dynamic algorithms, label propagation methods like dynamic overlapping SLPA (Xie et al. 2011) and other inherently dynamic and overlapping algorithms such as *AFOCS* (Nguyen et al. 2011). In the second main category, i.e., link communities, clustering is performed on links instead of nodes (Ahn et al. 2010). Here, link communities are mapped to node communities by finding nodes incidents to links within each community.

Further, there is an important class of soft clustering methods that utilize concepts specifically designed for vague and uncertain situations related to our proposed model. Decision-making under uncertain situation is identified as a basic concept underlying the human conception (Zadeh 1997) and also as a remarkable human ability to make rational decisions with partial knowledge. Yager and Filev (1998) believe that the "human has developed a granular view of the world" and "objects with which mankind perceives, measure and conceptualize and reasons are granular." This is totally consistent with the simplified representation of the real world by humans in which the precision in different levels of a social network is not a meaningful concept. Thus, the granulation concept is heavily used to present the situations involving uncertain and vague information. Here, instead of defining exact entities, the granules are defined. Granules are the clump of objects drawn together by some indiscernibility, functionality or similarity function. The inter-relation and intra-relation among granules are responsible for grouping

smaller granules into larger ones, or decomposing a large granule into smaller units. Any method in this line such as rough set and fuzzy sets is regarded as a subcategory of granular computing (Peters and Weber 2016) with different aggregation functions for the construction of the granules. Application of fuzzy methods in decision-making is already discussed in several papers (Fahmi et al. 2018; Oner and Oztaysi 2018). Fuzzy set theory addresses graded knowledge by fuzzy membership, and rough set theory defines knowledge granulation by the interdisciplinary relation and two sets of upper and lower bounds. A short description of these techniques is presented in the following definitions.

Definition 1 (*Fuzzy Set*). Let X be a set of objects. A fuzzy set A in X is a set of pairs $A = \{(x, \mu_A(x)) | x \in X\}$ where $\mu_A : X \rightarrow M$ is called the membership function of x in A mapping X into the membership space M ($M = [0, 1]$). Membership indicates the degree of similarity of an object x to an imprecise concept characterized by the fuzzy set A. The set of all elements having a positive membership in fuzzy set A constitutes its support set, i.e.,

$$\text{support}(A) = \{x | \mu_A(x) > 0\} \qquad (2)$$

Definition 2 (*Rough Set*). A granule in a rough set corresponds to a clump of objects drawn together by some indiscernibility, functionality or similarity function. Let E be such an equivalence relation on X. Any subset $R \subseteq X$ in the approximation space (X, E) is represented by its lower and upper approximations. The lower approximation $\underline{R}$ is the union of all the elementary sets that are subsets of X and the upper approximation $\overline{R}$ is the union of all the elementary sets that have a non-empty intersection with X. Finally, $\text{bnd}(R) = \overline{R} - \underline{R}$ is the boundary region of A where hesitant items lie; i.e.,

$$\underline{R}(X) = \{x | x \in V, [x]_R \subseteq X\} \qquad (3)$$

$$\overline{R}(X) = \{x | x \in V, [x]_R \cap X \neq \phi\} \qquad (4)$$

Information granules in fuzzy clustering arise by minimizing an objective function which expresses the spread of data $X = \{x_1, x_2, \ldots, x_n\}$ around prototypes $Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^m ||x_k - c_i||^2$, where c stands for the number of clusters. The clusters are described in terms of a family of prototypes $C = \{c_1, c_2, \ldots, c_c\}$. Numerous research fall in this category; however, most fuzzy-based methods require a priori knowledge about networks, e.g., the number of communities (Wang et al. 2013) or other fine-tuned parameters such as probability threshold (Breve and Zhao 2013).

Liu et al. (2015) proposed a granular-computing-based clustering algorithm where a granule is a subset of data with similar features according to their distances. For a complete review of methods in this category, please refer to Amelio

and Pizzuti (2014). Kundu and Pal (2015) suppose granule construction around a node with fuzzy boundaries. In this approach, granules are constructed around each node with a fuzzy boundary, which takes its membership degree according to the following equation:

$$\mu_c(v) = \begin{cases} 0 & \text{for } d(c,v) > r \\ \frac{1}{1+d(c,v)} & \text{otherwise} \end{cases}$$

where the distance measure $d(c,v)$ can be any metric, e.g., the weighted hop distance from node v to the center node c as mentioned in the paper. For the application of clustering, the authors find the granular embeddedness of all granules in the network, where this value for a pair of nodes a and b is defined as:

$$\varepsilon(a,b) = |A_a \cap A_b| = \sum_{v \in V} \min(\mu_a(v), \mu_b(v)) \qquad (5)$$

where A_a and A_a are the fuzzy sets representing the granules having the center nodes a and b. Then, it takes a hierarchical agglomerative approach and merges similar granules together. Dillen and Chakraborty (2016) used the FGSN framework to present their modularity-based community detection algorithm. However, the algorithm is designed for non-overlapping cases.

Granulation based on rough sets has found its application to clustering. In this approach, granules are chunks of objects drawn together by a similarity function. First, Lingras and West (2004) used the rough set concept in k-means clustering. Mean computation, assignment of objects to the approximations, and checking the stopping criterion are the three key features of rough k-means. The calculation of the modified centroid in rough k-means is given by the following equation:

$$c_i = \begin{cases} \omega_{\text{lower}} \times \frac{\sum_{x \in \underline{R}(c_i)} x}{|\underline{R}(c_i)|} + \omega_{\text{Upper}} \times \frac{\sum_{x \in \text{bnd}(c_i)} x}{|\text{bnd}(c_i)|} \end{cases} \qquad (6)$$

where $\omega_L + \omega_U = 1$ and ω_L and ω_U correspond to the relative importance of the lower and upper bound. For assigning the objects to the upper and lower bound of clusters, the ratio of $d(x,c_i)/d(x,c_j)$, $1 \le i, j \le k$ is used, where $d(x,c_i)$ is the distance between any object x and the centroid of cluster c_i. Suppose that $d(x,c_i) = \min_{1 \le j \le k} d(x,c_j)$ and $T = \{j : d(x,c_i)/d(x,c_j) \ge \text{threshold and } i \ne j\}$. If $T \ne \emptyset, x \in \overline{R}(c_j)$, x is not part of any lower bound. Otherwise, if $T = \emptyset, x \in \underline{R}(c_j)$. This algorithm and its derivation depend on the three parameters of ω_L, ω_U, and *threshold*. All the derivation of k-means using soft methods inherits the problems of k-means, including the requirement of passing the number of communities to the algorithm, and their sensitivity to initial seeds. Peters and Weber (2009) proposed a

preliminary idea of dynamic rough clustering based on rough k-means (Lingras and West 2004). The algorithm inherits the same challenges of rough k-means, i.e., determining the optimal number of clusters and approximation weights in each snapshot. Gupta et al. (2016) used rough set concepts in their clustering approach, where a granule is formed by using a neighborhood connectedness around each node. Then, a measure called relative connectedness of the neighborhood subset is calculated for each node and all the nodes having the same measure are merged together. A comparison of classical k-means and its fuzzy/rough version is discussed in Peters et al. (2013).

As the related work in this section explains, granular clustering realized using the fuzzy set and rough set suffer from the complexity and tuning of various parameter used in the model. In the following, we explain our work which is able to detect the number of communities automatically and has fewer parameters compared to rough clustering method.

3 Proposed model: rough granular social network community mining

Taking a human perspective of the social network in which granulation is used to perceive, measure and conceptualize objects in the world, we leverage the concept of granules in different levels of nodes and communities to design our adaptive community mining algorithm called "*Rough Granular Social Network Community Detection Approach.*" The structure of the macrogranules incrementally emerges based on the interaction of close- enough microgranules which are joined together over time to construct the higher level modules of the network. This process involving vagueness and rough concepts is used to model this uncertainty. The following section describes some notation and definitions which will be needed later to describe the mechanism of the construction of the final macrogranules in temporal networks.

3.1 RGSN: model description

Let us consider a temporal weighted/unweighted network by several static snapshots of the network $\Delta = (G^1, G^2, \ldots, G^s)$. A snapshot of the network is represented by $G^t = (X^t, E^t)$ where X^t is the set of nodes available in time step t and $E^t \in (X^t \times X^t)$ is the set of edges available. The number of nodes may be changed in different time steps allowing for insertion and deletion of nodes. The interaction of nodes with each other can be represented in an adjacency matrix $W^t = (w_{ij}^t)_{n \times n}$ where $w_{ij}^t = 1$ denotes there is an edge between node i and j and other values denotes their weight of interaction. We denote the members of the network in time step t as $X^t = \{x_1, \ldots, x_n\}$ where each member includes its attribute including node id (v_i), *timestamp* (t)

and *weight* of the interaction with their neighborhood $\Gamma(x_i)$ where $\Gamma(x_i) = \{x_j \in X | (x_i, x_j) \in E\}$, i.e., (w_{ij}). Further, $\Omega = \{C_1, \ldots, C_k\}$ is a family of clusters classifying objects into k communities.

A social network as an instance of a complex system can be characterized by the interaction that occurs between two levels of the network. From this viewpoint, the interaction among the basic microcomponents of the network, i.e., the nodes, allows for the creation of the some macrostructure of the network, i.e., communities, at the higher level of the system. This micro–macro mechanism is heavily discussed in emergence theories and third wave of system theories, which are well-linked to social sciences (Sawyer 2005). Inspired by this mechanism, we model the network at both levels of micro- and macrostructures as granules. Let a microgranule be associated with each node, i.e., $x_i \in X$ be $\varphi(x_i)$. In addition, the final community structure be represented by macrogranule structure Ω. Now, we model the network by a rough granular framework represented by a pair $S = (\varphi, \Omega)$ where φ is a finite set of rough microgranules around each node, i.e., $\varphi = \{\cup \varphi(x_i) | x_i \in X\}$, and Ω is a finite set of rough macrogranules constructed over time by the interaction of close-enough microgranules. In the following section, we present some definitions used in the proposed model.

3.1.1 Rough microgranule

Today, the importance of the environment within which each individual lives has a tremendous effect on his/her decision-making process. This is long discussed in different modern social theories and is observed as de-facto. Taking this viewpoint, we model each individual as a microgranule represented together with his/her direct neighborhood. To realize the granular structure of each individual, the rough set concept is used. In the rough terminology, the members of the set are splits into two parts, deterministic and indeterministic, drawn together by similarity function. Thus, each center node is placed in the lower approximation of microgranules, and other neighbors are associated with the granule by a variation of similarity which is more highlighted if the network is weighted (neighbors with higher volumes of interaction are more similar to the center node). Hence, we give the following definition for *rough microgranule*:

Definition 1 (*Rough microgranule*). A microgranule around a center node x_i and annotated by $\varphi(x_i)$ consists of the attributes for the center node x_i and its direct neighborhood class $\Gamma(x_i)$. Thus, the network can be represented by a set of microgranules constructed around each node $(G^t = \{\varphi(x_i), \ldots, \varphi(x_n)\}$ for all $v_i \in V)$. The attribute of each node in the granule is related to the network properties

of these nodes. We have specified the following attributes for each member of the microgranules:

$$\varphi(x_i) = \begin{cases} \varphi_V(x_i) = \{v_i \cup v_j | (v_i, v_j) \in E\}, \\ \varphi_W(x_i) = \left\{\sum w_{ij} \cup w_{ij} | j \in \Gamma(x_i)\right\}, \\ \varphi_T(x_i) = t, \\ \varphi_\mu(x_i) = \left\{1 \cup \dfrac{w_{ij}}{\sum_{j \in \Gamma(x_i)} w_{ij}}\right\} \end{cases} \quad (7)$$

where

- $\varphi_V(x_i)$ stores the vertex vector (V) comprised of the node ids for the center node x_i and its neighbors;
- $\varphi_W(x_i)$ stores the weight vector (W) for each member of vector V. For the neighbor node $j \in \Gamma(x_i)$, this is equal to the weight of each neighbor j connected to the center node (w_{ij}) and for the center node (x_i), this will be the sum of weights $\sum w_{ij}$.
- $\varphi_T(x_i)$ records the time step that the nodes in the vector V are observed.
- Finally, $\varphi_\mu(x_i)$ stores the *participation* vector (μ) for each member of vector V derived by the calculation of the weight value normalized by the sum of weights of the center node. The center node gets the highest participation point of 1, and other members acquire some value less than one proportional to their weight, i.e., $\frac{w_{ij}}{\sum_{j \in \Gamma(x_i)} w_{ij}}$. The attribute is similar to the membership degree in fuzzy set, and is an implication of the belongingness of each member to the microgranule. The graphical representation of this structure is illustrated in Fig. 1.

3.1.2 Rough macrogranule

Naturally, each community may hold members shared by other communities producing an overlapping structure. The belongingness degree of individuals to different communities varies. Hence, we model each community in the framework of rough sets. In this model, the members which are determined *surely* to belong to a community constitute the lower approximation of the rough macrogranule intended to represent that community, and the members in the overlapping region with other communities constitute its upper approximation.

Definition 3 (*Rough macrogranule*). Let $\Omega^t(l) (l = 1, \ldots, k)$ be the l'th community discovered in time step t and the lower and upper approximation of this community be $\underline{\Omega}_l^t$ and $\overline{\Omega}_l^t$, respectively. The *lower approximation of the rough macrogranule* set are the members which are determined to be definitely in the community Ω_l. On the other hand, the upper approximation set of the macrogranule

Fig. 1 Microgranule structure of vertex x_i
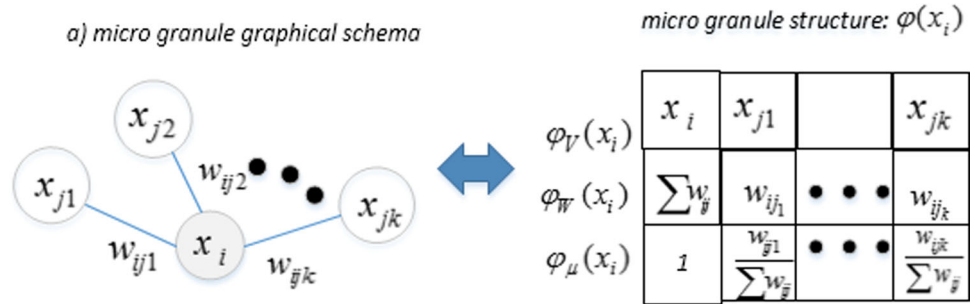
l ($\overline{\Omega}_l$) is constituted of the members which are *probably* inside the community. These members are in the overlapping regions with other communities, and some of them may later be added to the lower approximation of the microgranule l ($\underline{\Omega}_l$). This is mathematically expressed in the following equations:

$$\underline{\Omega}^t(l) = \{x_j | x_j \in \text{Support}(\Omega(x_p) \wedge x_j \in \text{Support}(\varphi(x_q));$$
$$\forall \Omega(x_p) \in \Omega_i \text{ and } \forall \varphi(x_q) \in \Omega_j; i \neq j\} \quad (8)$$

$$\overline{\Omega}^t(l) = \{x | x \in \text{Support}(\varphi(x_p)); \forall \varphi(x_p) \in \Omega_i\}. \quad (9)$$

Both the lower and the upper approximation structure of the macrogranules have similar attributes to the microgranules, i.e., $\overline{\Omega}^t(l) : \{\overline{\Omega}^t_V(l), \overline{\Omega}^t_W(l), \overline{\Omega}^t_\mu(l)\}$. The values of

these attributes are updated from the microgranule structures joined into the macrogranule structure. A detailed explanation of each attribute is described in the update process.

3.2 Algorithm description

In this section, we outline how the final communities comprised of densely connected nodes are formed in temporal context. Exploiting the granular model of social network as stated in Sect. 3.1, we first present a general view of the algorithm in a short step-by-step schema. Next, we provide a detailed explanation of different steps. The high-level procedure of rough community detection method is represented in Pseudo code 1.

Pseudo-code 1. High level description of the proposed algorithm

Input: Graph at t^{th} snapshot (G^t) or streaming data

Step 1. *Granule Initialization*

　　a. Initialization of macro granules (prototypes) using the previous macro granules (for $t > 1$)

　　b. Initialization of the next micro granule by receiving new data entry and its available neighborhood information (Eq. 8 & Eq. 9)

Step 2. *Similarity calculation*

　　a. Calculate the rough granule embeddedness of the micro granule into available macro granules (Eq. 10–12)

　　b. Find the best candidate macro granule by the calculation of the maximum weighted rough granule embeddedness (Eq. 13)

Step 3. *Update community structure*

Step 4. *Termination or Go back to Step 1.*

　　a. At the end of each time step, output the final disjoint/overlapping macro granules (communities)


Granular-ARTISON models the networks into granule structure in both the node (micro) and the community (macro) level, similar to human granular perception way of perceiving nodes and communities in the network. Next, the microgranules join the macrogranule structure which they find the most *similar*. Group formation based on some similarity measure is verified in different studies. This is the way that is already followed by the representative or partitioning-based algorithms. If there is not any similarly *enough* macrogranule to join, a new one will be created. The process of the similarity determination of each microgranule with the available macrogranules is done through a two-step process, which comprises of first selecting some candidate communities (Eq. 11) and then refining the selection by finding the best match (Eq. 14). The two-step process of determination is similar to the way humans are involved in the decision-making process (Grossberg 2013). The uncertainty is observed in different stages of network modeling and decision-making. For this reason, we incorporate the uncertainty in different parts of the algorithm, specifically in the modeling of the nodes and communities and the similarity measure designed. The structure and nodes in the macrogranules are modified and tuned-based on the interactions of nodes observed. In the end, the most probable members of each community that have dense interactions with each other compared to other members will be stored in the lower approximation of each macrogranule, and the other fringe items with fewer interaction are stored in the upper approximation of the macrogranule. Hence, there may be members with different degree of memberships to a community. For the purpose of disjoint communities, only the members with high probability of belongingness to each community are considered (lower approximation members of each macrogranule) and for the overlapping case, the members present in both the lower and upper approximation are included. The members with high participation ratio are considered as the core nodes, and the other members in upper approximation construct the overlapping region of each community. Throughout these processes, the number of communities is gradually found based on the interaction information. Having described a total schema of the algorithm, we go through a detailed explanation in the following sections. An illustration of the algorithm's components including microgranule and macrogranule in two levels of node and community is depicted in Fig. 2 and will be discussed in the following parts.

3.2.1 Initialization

First, the basic elements of the framework, i.e., the microgranules, are constructed iteratively according to the information available (based on Eq. 6). After each microgranule is created, it searches for the similar macrogranules (communities) in which to be included, as discussed in the next
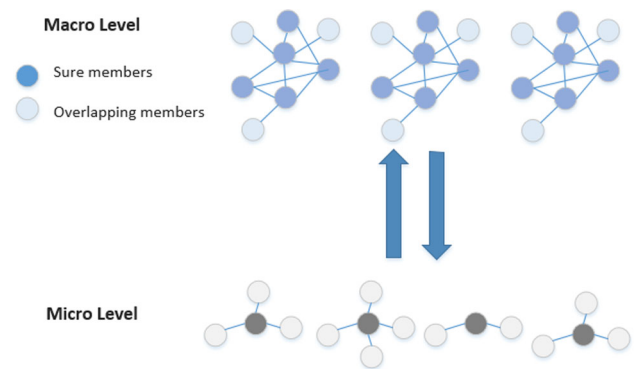
Fig. 2 A schema of the interactions observed in Granular-ARTISON algorithm in two micro- and macro-level (microgranule observed in the node level leads to the formation of macrogranules which are adapted in different time steps)
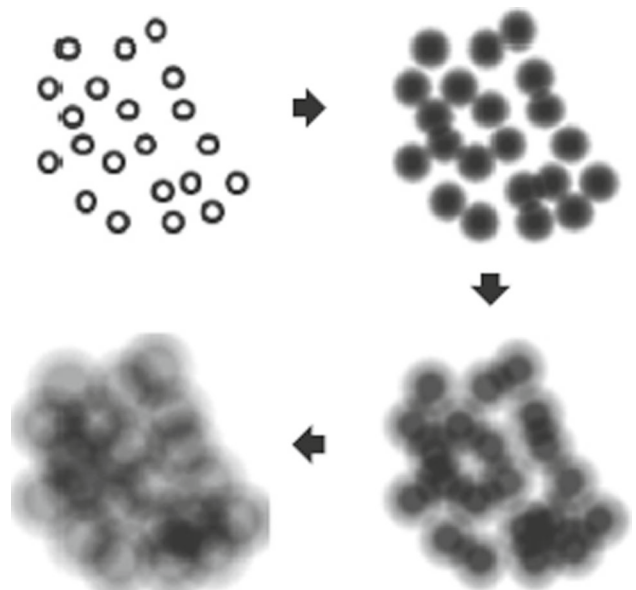

Fig. 3 Node granulation concept using circle of connections

step. Further, at the beginning of each time step, the community prototypes are initialized with the rough macrogranules obtained in the previous time step to preserve the temporal smoothness. Figure 3 shows the network structure composed of microgranules.

3.2.2 Rough granular similarity calculation

Humans are always intertwined with decision-making in uncertain situations. Here, we define a novel similarity measure that directly integrates the uncertainty of the decision-making process in its formulation. In this way, we favor the similarity obtained through sure level of rough set compared to similarity obtained by unsure level. The similarity calculation process is illustrated in the following.

Fig. 4 An example of the construction of the first macrogranule based on the first microgranule. **a** Microgranule graphical schema. **b** Microgranule structure: $\varphi(1)$. **c** Lower macrostructure: $\underline{\mathbb{C}}(1)$. **d** Upper macrostructure: $\overline{\mathbb{C}}(1)$
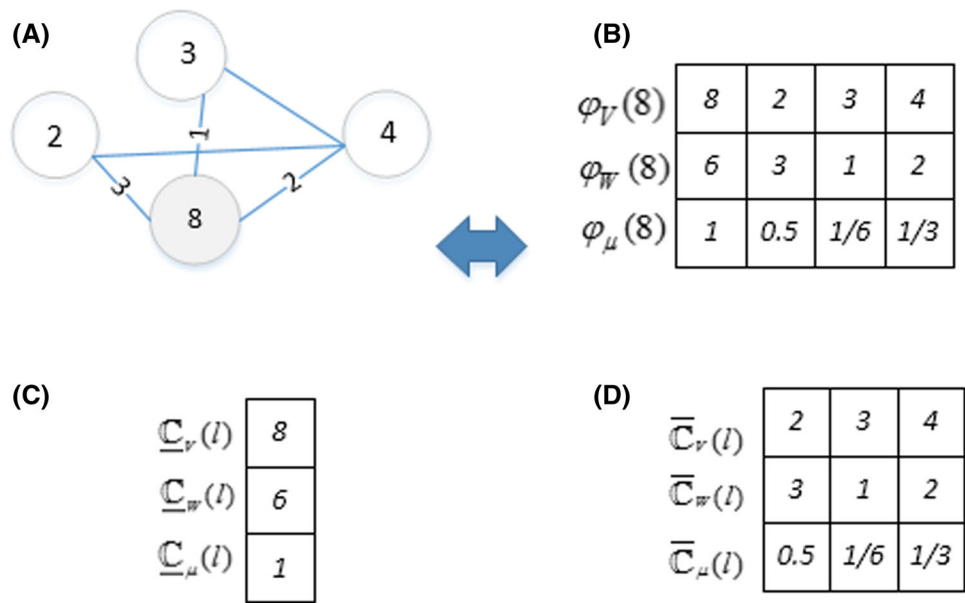
First, we will determine for each node x_i under examination, along with its neighbors $\Gamma(x_i)$ constructing the microgranule $\varphi(x_i)$, the most similar macrogranules $\Omega^t(l)(l = 1, \ldots, k)$ for incorporating the microgranule. Obviously, at the beginning of the first timestamp, there is no macrogranule available and the first macrogranule ($\Omega(1)$) is constructed based on the microgranule $\varphi(x_i)$. In this case, the center node x_i is placed in the lower approximation set of the macrogranule, since we are in search of the most similar macrogranule for this center member and each member is *surely* a subset of itself. The other neighbors of the center node ($\Gamma(x_i)$) are *probably* members of this community and are placed in the upper approximation of this macrogranule. They resemble some similarity with the sure member of the community (the center node) through the neighborhood relation. An example of microgranule is schematically represented in Fig. 4.

The appropriate similarity measure should be used to assess the similarity of each microgranule against multiple macrogranules. Notice that the macrogranules are presented in rough concepts, i.e., some members *surely* belong to the macrogranule (placed in the lower approximation), and some others *probably* belong to this structure (upper approximation). Hence, it is rational to consider a *rough granular weighted similarity* measure to differentiate between the similarity values obtained from these two approximations and assign a higher weight to the term that results from the comparison with certain elements of the macrogranules. The *rough granular weighted similarity* measure is formally defined as follows:

Definition 4 For two given microgranule $\varphi(x_i)$ and macrogranule Ω_q, the *rough granular participation ratio* denoted as $sim^\mu(\varphi(x_i), \Omega_q)$ is a weighted additive measure composed of the *rough granular lower participation ratio* and

rough granular upper participation ratio measure linked by two different weights of (α, β), where $\alpha > \beta$; i.e.,

$$sim^\mu(\varphi(x_i), \Omega(q)) = \alpha * \underline{sim}^\mu(\varphi(x_i), \underline{\Omega}(q)) \\ + \beta * \overline{sim}^\mu(\varphi(x_i), \overline{\Omega}(q)). \quad (10)$$

These two terms are derived from the similarity value of the microgranule $\varphi(x_i)$ in both the lower and the upper approximation of the macrogranule $\Omega(q)$. How these two similarity measures are calculated is explained as follows.

The *rough granular lower (upper) participation ratio* measure for a given microgranule $\varphi(x_i)$ and macrogranule $\Omega(q)$ is defined as the ratio of the participation degree of the microgranule $\varphi(x_i)$ into the lower (upper) approximation of the macrogranule $\Omega(q)$. This is calculated by the intersection of the participation ratio of the microgranule and the lower (upper) macrogranule normalized by the value of this attribute in the microgranule, i.e.,

$$\underline{sim}^\mu(\varphi(x_i), \underline{\Omega}(q)) = \frac{\sum_{j \in \varphi_v(x_i) \cap \underline{\Omega}_v(q)} \varphi_{\mu_j}(x_i) \cap \underline{\Omega}_{\mu_j}(q)}{\sum \varphi_\mu(x_i)} \quad (11)$$

$$\overline{sim}^\mu(\varphi(x_i), \overline{\Omega}_q) = \frac{\sum_{j \in \varphi_\mu(x_i) \cap \bar{\Omega}_q(v)} \varphi_{\mu_j}(x_i) \cap \bar{\Omega}_{\mu_j}(q)}{\sum \varphi_\mu(x_i)} \quad (12)$$

The microgranule is checked against all available macrogranule structures to find the prototypes which have *rough granule participation ratio* higher than a threshold value. The threshold value is a minimal similarity level which is required for the inclusion of a microgranule into higher macrogranule

structure. Obviously, if there is not any macrogranule satisfying the threshold criterion, a new community should be created to incorporate the microgranule. Otherwise, if there are several communities higher than the threshold value, the community which has the highest value of similarity is chosen to incorporate the microgranule. The decision-making process is translated into the following relation:

$$\Omega^* = \max(sim^\mu(\varphi(x_p), \Omega(q)) > \partial) \tag{13}$$

In this way, the community that is most similar in terms of both activeness and embeddedness is selected to incorporate the microgranule members. We have chosen the threshold value of 0.3 which gives good results in various datasets.

3.2.3 Updating scheme

After the selection of the macrogranule, which should integrate the microgranule members, an update scheme should take place to account for this assignment. The core node of the microgranule is added to the lower approximation of the selected macrogranule. Further, the neighbors of the microgranule are added to the upper approximation of this macrogranule. As described in Sect. 3.1.2, the macrogranule structure has four fields of *ids, weights, timestamp, and participation ratio* in both the lower and the upper approximations $\Omega(l) : \{\underline{\Omega}(l), \overline{\Omega}(l)\}$. Here, we give an explanation of the update process in these fields related to the lower approximation $\underline{\Omega}^t(l) : \{\underline{\Omega}_V^t(l), \underline{\Omega}_W^t(l), \underline{\Omega}_\mu^t(l)\}$:

- *Id* attribute: $(\underline{\Omega}_V^t(l))$: this is the identity of each member present in the lower approximation of the macrogranule l. The other attributes are defined for each member of this set (V). In each update process, the center node is added to the lower approximation of the macrogranule, and the neighbors are added to the upper approximation (if not already present).
- *Weight* $(\underline{\Omega}_W^t(l))$: it indicates the sum of the weighted interactions observed during different iterations of the algorithm. The higher the weighted sum of a member is, the more weighted connection it has till the present time and is a better representative for that community. For the nodes added to the lower or upper approximation of the set, this attribute will be updated by adding previous values and the values of the newly added microgranule members. The final maximum value for each member will be the sum of weights of the member in the network.
- *Normalized participation ratio* $(\Omega_\mu^t(l))$: the attribute indicates the participation grade of the microgranule members in the macrogranule structure. The values of this attribute get updated using the rough granular similarity measure obtained for the higher the value of this attribute for a member, the higher is the interaction level of this macrogranule

in the; i.e., the node has the higher number of connections with the other members of the community.

This community assignment scheme allows for an overlapping community structure. The certain members are in the lower approximation of the macrogranule and overlapping members in the upper approximation.

4 Experimental results

In this section, we represent the results of the experiments carried out to assess the performance of the algorithm on several evolving synthetic (Greene et al. 2010), and real datasets (Zachary 1977; Lusseau et al. 2003; Steinhaeuser and Chawla 2008; Lin et al. 2008). First, we describe different aspects of evaluation including the introduction of datasets, algorithms and measures used. Then, the results of the experiments on synthetic and real datasets are investigated.

4.1 Experiment setup

4.1.1 Datasets

For synthetic dataset generation, we use the scenarios generated by the dynamic version of the frequently used LFR generator (Greene et al. 2010), which is one of the best generators producing similar properties to real networks. The availability of ground truth information makes these networks especially interesting for evaluation purposes. Each experiment is executed for 50 time, and average values are reported. Further, several famous benchmarks including Zachary Karate club (Zachary 1977), Doubtful Sound Dolphins (Lusseau et al. 2003), and the Risk Game network (Steinhaeuser and Chawla 2010) are used to evaluate the accuracy of the algorithm. Finally, the performance of the proposed algorithm in the evolving network dataset is also examined by the NEC blog dataset (Lin et al. 2008). Our programming environment is MATLAB.

4.1.2 Compared algorithms

Granular-ARTISON is placed in the representative-based algorithm class, the best comparison is achieved by comparing it to other representative-based algorithms. For this reason, we choose two other evolutionary representative-based algorithms especially designed for dynamic settings. We use the state-of-the-art evolutionary framework called Adaptive Evolutionary Clustering (AFFECT (Xu et al. 2014)), extended to the k-means and spectral algorithm, in which the optimal smoothing factor is determined automatically using a statistical approach. Both algorithms require an estimation of the number of communities for each timestamp. For this purpose, we use the well-known silhouette width criterion

Table 1 Comparison of proposed algorithm in synthetic dataset with AFFECT k-means and AFFECT spectral in mean values in five measures of (a) Rand Index, (b) NMI, (c) F1, (d) modularity and (e) accuracy of the no. of clusters

Network event	Measure	Granular-ARTISON	AFFECT k-means	AFFECT spectral
Birth and death event	Rand Index	**0.92** (± 0.03)	0.46 (± 0.39)	0. 87 (± 0.01)
	NMI	**0.63** (± 0.02)	0.31 (± 0.23)	0.55 (± 0.01)
	F-measure	**0.57** (± 0.03)	0.25 (± 0.21)	0.47 (± 0.01)
	Modularity	0.61 (± 0.09)	0.32 (± 0.29)	**0.70** (± 0.01)
	No. of cluster accuracy	**0.51** (± 0.01)	0.50 (± 0.01)	0.50 (± 0.01)

for automatic determination of the number of communities. This measure determines how compact the distance of communities is for a given time step. The maximum width of this measure is used to assess the number of needed communities in k-means. Since both algorithms are random-based, they produce different results in experimental runs. Hence, the average value of different measures in these two algorithms is reported in the experiments. Finally, we use two other methods in real dataset evaluation section. These algorithms are in the category of modularity-based and label propagation-based algorithms called Louvain (Blondel et al. 2008) and SLPA (Xie et al. 2011). Both algorithms show high accuracy and are selected as representative algorithms in their category. We use the threshold value of $\delta = 0.3$, which yield good results using experiments.

For the evaluation of the proposed algorithm in overlapping case, two highly used algorithms in overlapping community detection domain with available code are selected. The first is Link-Clustering (Ahn et al. 2010) algorithm as the pioneer link-based algorithm and the second is OCG (Becker et al. 2011) which has several similar functionalities to Granular-ARTISON. Specifically, OCG also uses a local similarity-based algorithm where different segments of the network are merged into each other to construct communities.

4.1.3 Compared measures

The standard approach for assessing the accuracy of community detection results is to exploit the external measures that compare the accuracy of detected communities with ground truth communities. We use a range of measures including the Rand Index, NMI and the F-measure to determine the accuracy of the community detection algorithms in different scenarios, as will be explained. The NMI measure indicates how much knowing one of these communities helps predict the structure of the other and reduce prediction uncertainty. This measure has proven to be a robust and accurate similarity measure for many modalities. It is bounded in [0, 1] to verify the complete sharing of the partitions found when the value is equal to 1 and the complete dependence of the partitions

when it equals 0. The F-measure presents a harmonic means of precision and recall measures, where precision is the ratio of relevant objects (real community members detected) to the total number of objects detected, and recall is the ratio of relevant objects detected to the total number of objects based on ground truth. All these measures reach their best at 1 and their worst at zero value. Further, the internal measure of modularity is used to assess the quality of the communities in real networks. We use the well-known modularity measure for this purpose. In addition, when the functioning of the algorithm in overlapping case is assessed, the overlapping-NMI (McDaid et al. 2011) specifically measure is used which is designed to specifically assess the accuracy of clustering in overlapping case. Another measure in this context is the accuracy of the number of clusters each algorithm is able to detect which is used in the evaluation of algorithms in this category.

4.2 Artificial networks

To evaluate the performance of the algorithm in dynamic settings, it is important to assess the algorithm on dynamic networks where ground truth information is available. Dynamic network environments incorporate different events of birth, merge and expansion/contraction sections. For this purpose, we use the dynamic synthetic dataset generator introduced by Greene et al. (2010), which is derived from the well-known LFR benchmark network. The benchmark can create different events in dynamic scenarios and inherits the basic statistical properties of the real networks in heterogeneous distributions of the degree and the community size. Different parameters of this benchmark are tunable, which allows for overlapping and dynamic network settings. These parameters include: size of the network N, size of the communities (within $C_{\min}$ to $C_{\max}$), and mixing parameter, i.e., the overlap among communities (μ). The combination of these settings helps to analyze the algorithm in detail.

In our experiments, 1000 nodes in five time steps undergo different events to evaluate the performance of the algorithm. The number of nodes may change during different events. The generated scenarios follow standard settings for produc-
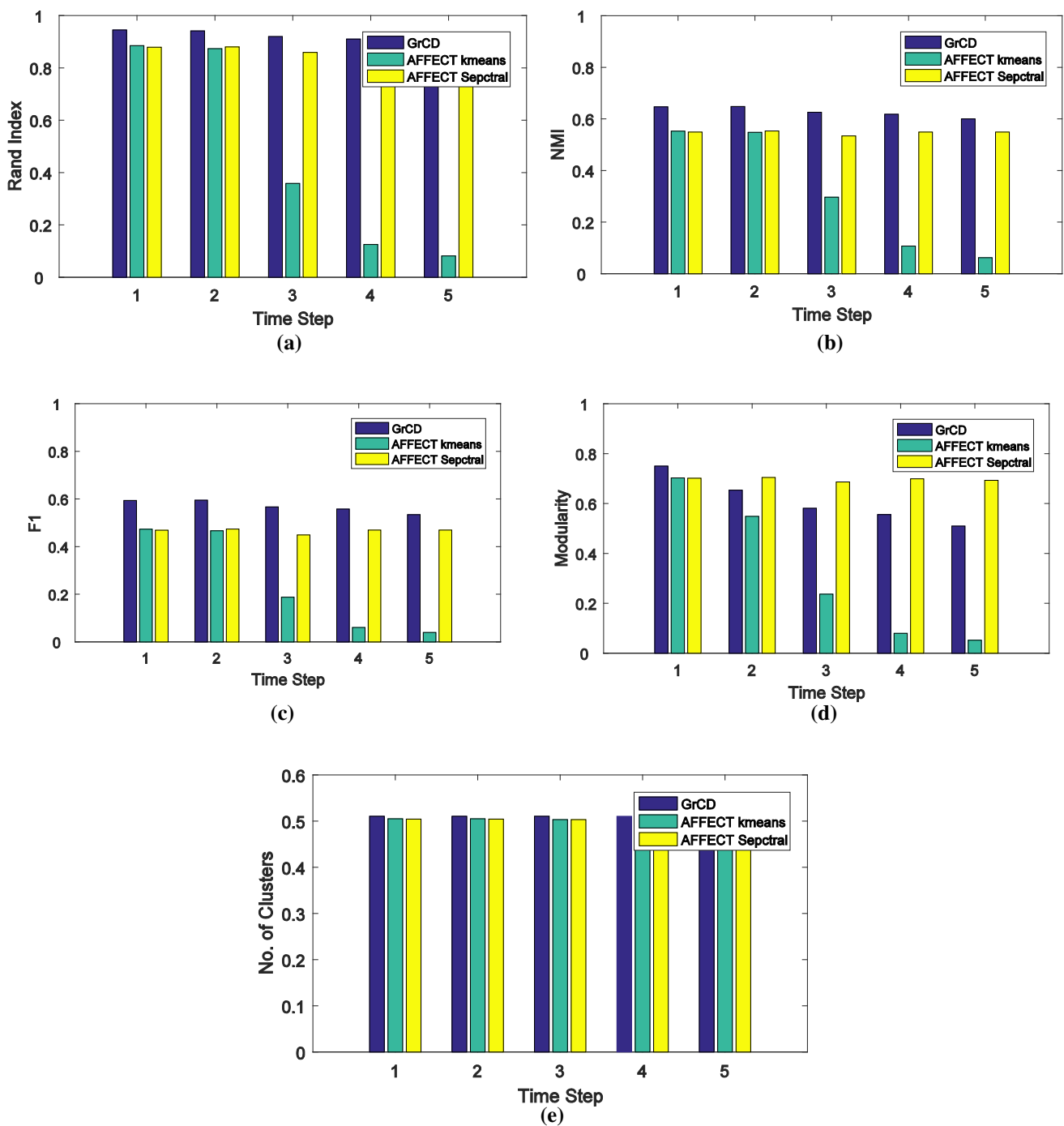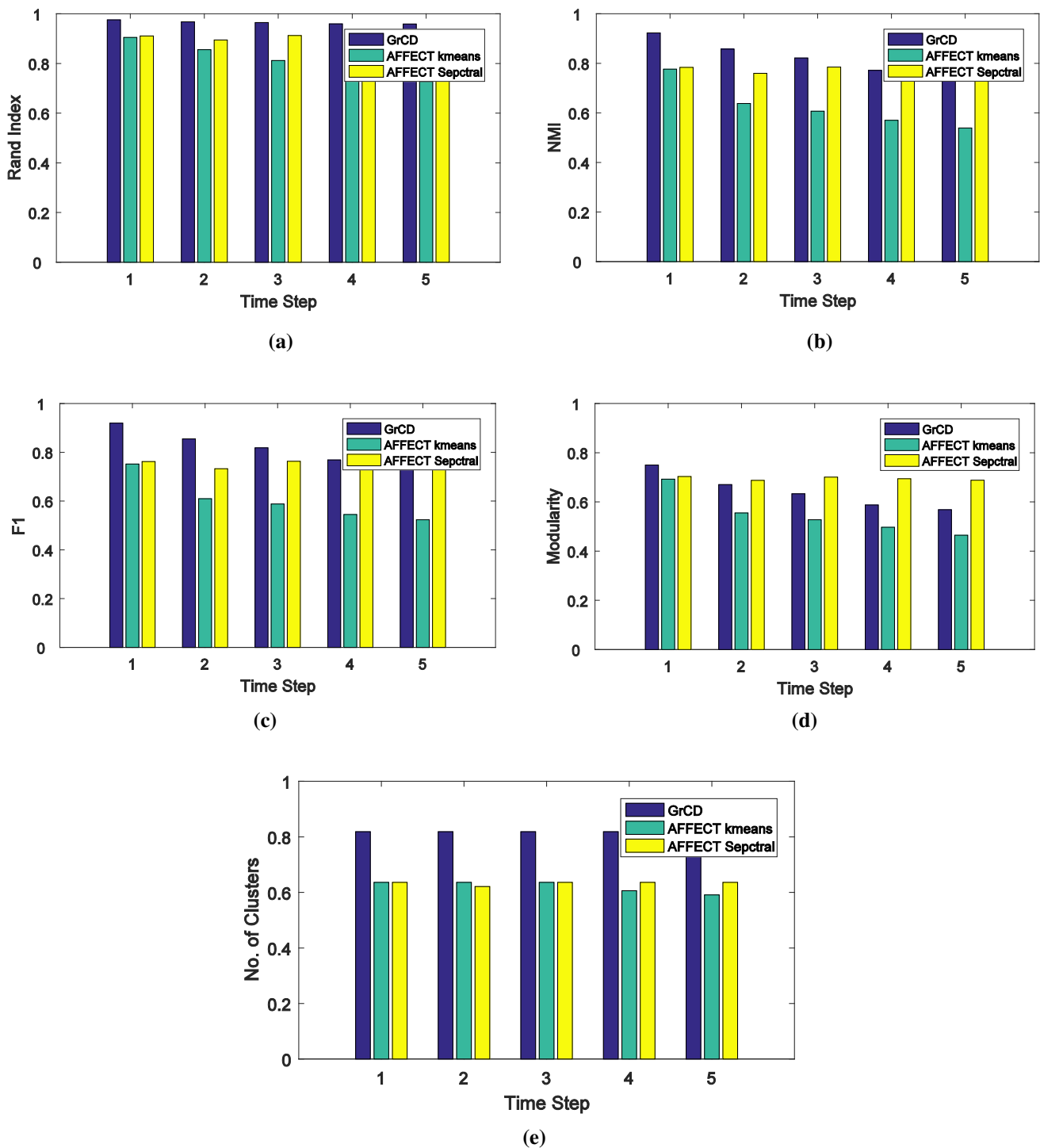
Fig. 5 Performance comparison of Granular-ARTISON in the *birth and death* event of the synthetic data generator with AFFECT k-means and AFFECT spectral in five measures of **a** Rand Index, **b** NMI, **c** F1, **d** modularity and **e** accuracy of the no. of clusters

ing power-low networks as used in Folino and Pizzuti (2014). Events used in the experiments are: (1) birth and death; (2) expansion and contraction; (3) merging and splitting and (4) intermittent communities. The average and maximum degree is set to 20 and 50, respectively, and the minimum and maximum community size is 10 and 50 nodes, respectively. The

initial number of nodes is 1000 nodes, and the experiments are averaged over repeated runs for consistency.

4.2.1 Birth and death event

In this experiment, 10% of new communities are created by removing nodes from other existing communities, and

Fig. 6 Performance comparison of Granular-ARTISON in the *expansion and contraction* event of the synthetic data generator with AFFECT k-means and AFFECT spectral in five measures of **a** Rand Index, **b** NMI, **c** F1, **d** modularity and **e** accuracy of the no. of clusters

randomly removing 10% of the existing communities. The results of the experiments are averaged in Table 1. The values of measures obtained in each time step are also illustrated in Fig. 5. The number of nodes decreases from 1000 nodes in the beginning to 784 nodes in the last run of the algorithm.

This event is one of the hardest events to capture for all algorithms. As illustrated in Fig. 5, the performance is degraded in the NMI, F-measure and modularity measures. AFFECT k-means encounters serious problems from the third time step, and the performance in all measures dete-

Table 2 Comparison of proposed algorithm in synthetic dataset with AFFECT k-means and AFFECT spectral in five measures of (a) Rand Index, (b) NMI, (c) F1, (d) modularity and (e) accuracy of the no. of clusters

Network event	Measure	Granular-ARTISON	AFFECT k-means	AFFECT spectral
Expansion and contraction event	Rand Index	**0.97** ($\pm$0.01)	0.81 ($\pm$0.07)	0.91 ($\pm$0)
	NMI	**0.83** ($\pm$0.05)	0.63 ($\pm$0.05)	0.78 ($\pm$0.04)
	F-measure	**0.82** ($\pm$0.07)	0.60 ($\pm$0.09)	0.75 ($\pm$0.01)
	Modularity	0.64 ($\pm$0.07)	0.55 ($\pm$0.09)	**0.69** ($\pm$0.01)
	No. of cluster accuracy	**0.82** ($\pm$0.0)	0.62 ($\pm$0.02)	0.63 ($\pm$0.01)

riorates. Meanwhile, Granular-ARTISON preserves high differences with AFFECT k-means in all measures by an average of 30%. Specifically, there is a 46% difference in rand index, which measures correct assignment of nodes to their communities. The NMI measure in Granular-ARTISON shows 32% and 8% improvement over AFFECT k-means and AFFECT spectral, respectively. The priority is kept on F-measures, too. However, the modularity of the AFFECT spectral is the only case where this algorithm stands higher than Granular-ARTISON. The higher value of the Rand index in Granular-ARTISON, which is the manifestation of the correct assignment of the nodes to their communities, compensates for this shortcoming. The number of clusters assessed by all algorithms is similar.

4.2.2 Expansion and contraction event

In this event, 10% of communities are randomly selected and expanded or contracted by 25% of their size. The number of nodes varies from 1000 in the beginning to 970 in the last run. The average values of measures are reported in Table 2, and the results of each time step are illustrated in Fig. 6. The average value of all measures in Granular-ARTISON is higher than the rest. The Rand index values of all algorithms are close to each other. AFFECT spectral stands in second place and AFFECT k-means in last place. The average values of NMI and F-measure in Granular-ARTISON show 20% and 8% improvement over AFFECT k-means and AFECT spectral, respectively. Like the birth and death event, the modularity value of AFFECT spectral is slightly higher than Granular-ARTISON. Overall, it appears that the expansion scenario is easier to capture in all algorithms.

4.2.3 Intermittent communities event

In this event, 10% of communities from the first time step hide. The results of this experiment are shown in Fig. 7, and the average value and standard deviations are reported in Table 3.

The number of nodes in time steps is as follows: 1000, 892, 917, 909 and 927 nodes. In this experiment, Granular-ARTISON is superior in all measures of the Rand Index, NMI, F-measure and modularity by almost 20% and 10% with respect to AFFECT k-means and AFFECT spectral, respectively. Further, the number of correctly guessed communities is also higher or equal to other algorithms in different steps (Fig. 7e).

4.2.4 Merging and splitting event

In merging and splitting events, 10% of communities are split, 10% of communities are chosen, and a couple of communities are merged at each time step. The average values of measures are reported in Table 4, and the results of each time step are illustrated in Fig. 8. The number of nodes is fixed, but various mergers and splits make the scenarios rather difficult for evolutionary-based algorithms. However, Granular-ARTISON results show its high performance compared with the other two algorithms by almost 20% in NMI, F-measure and modularity measures. Finally, the number of communities derived by Granular-ARTISON is closer to ground truth by achieving a mean accuracy of 83% compared to almost 60% of evolutionary-based algorithms.

We performed a two-tailed t test at a 0.05 level of significance tests were performed after ensuring that the data followed a normal distribution (by using the Kolmogorov–Smirnov test). The result of tests performed between the average value of F-measure for Granular-ARTISON and the two other evolutionary algorithms are reported in Table 5. Results indicate that Granular-ARTISON preserves its difference with other algorithms in all events ($<$0.05). Hence, the good performance of Granular-ARTISON compared to other algorithm can be expected in other datasets, too.

4.2.5 Overlapping scenarios

When Granular-ARTISON is working in the overlapping case, the upper bound members of each community are considered as its overlapping members. In this case, the performance of the algorithm is benchmarked against the other overlapping algorithms. As the results in Fig. 9 illustrates, Granular-ARTISON shows much better performance than
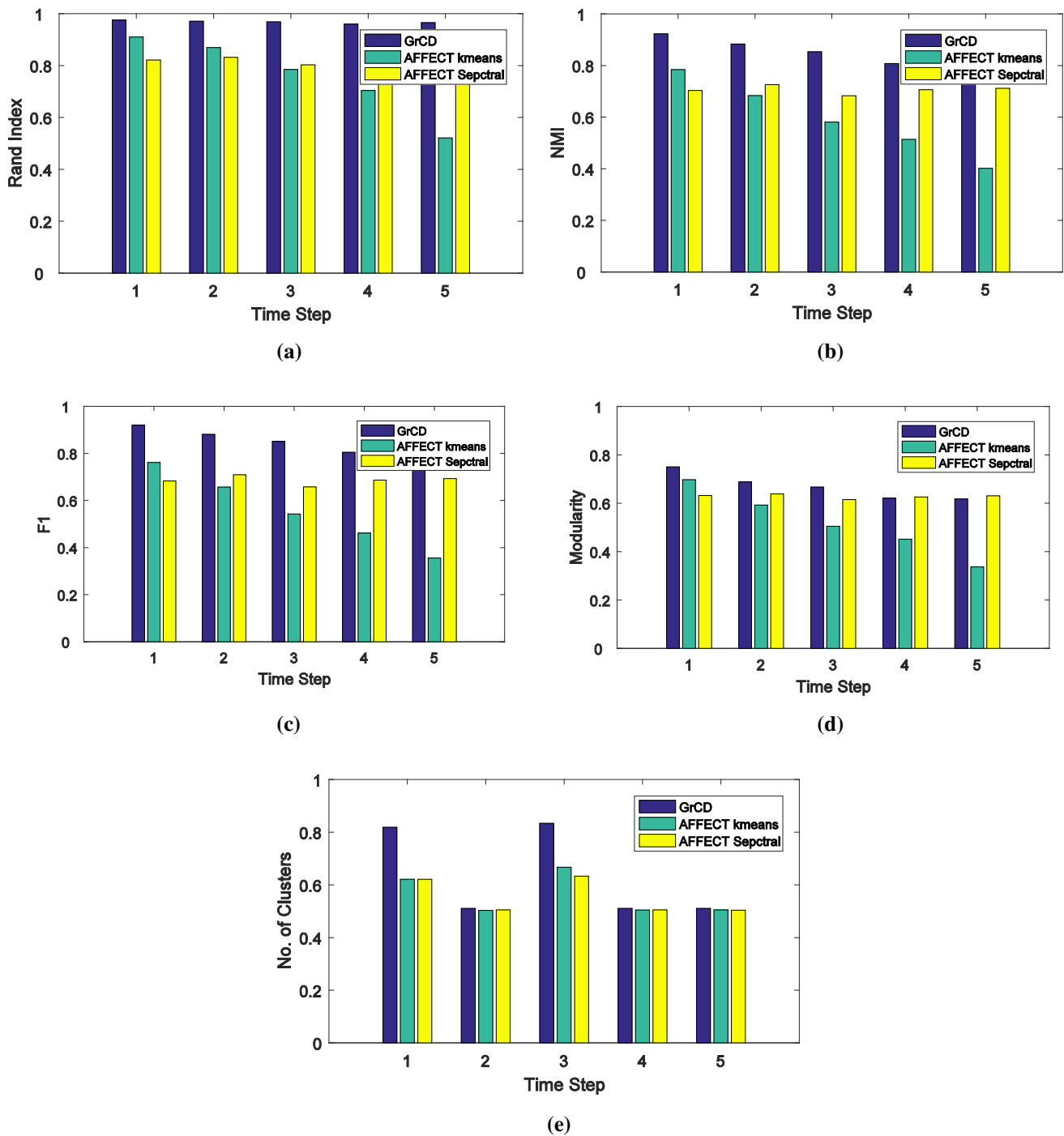
Fig. 7 Performance comparison of Granular-ARTISON in the *hiding* event of the synthetic data generator with AFFECT k-means and AFFECT spectral in five measures of **a** Rand Index, **b** NMI, **c** F, **d** modularity and **e** accuracy of the no. of clusters

the other two overlapping algorithms in overlapping-NMI measure.

In the best case of birth and death event, the algorithm on average shows 24% and 34% higher value than Link-Clustering and OCG algorithms, respectively. In all events Link-Clustering algorithm stands in the second place and the

lowest overlapping-NMI value is for OCG algorithm. The average and deviation values are reported in Table 6.

Another measure used to assess the accuracy of the algorithms in the number of clusters found is to investigate the number of clusters found by each algorithm and compare the value to the ground truth value. Figure 10 illustrates the

Table 3 Comparison of proposed algorithm in synthetic dataset with AFFECT k-means and AFFECT spectral in five measures of (a) Rand Index, (b) NMI, (c) F, (d) modularity and (e) accuracy of the no. of clusters

Network event	Measure	Granular-ARTISON	AFFECT k-means	AFFECT spectral
Birth and death	Rand Index	**0.97** ($\pm$0.01)	0.76 ($\pm$0.15)	0.82 ($\pm$0.01)
	NMI	**0.86** ($\pm$0.05)	0.59 ($\pm$0.15)	0.71 ($\pm$0.02)
	F-measure	**0.85** ($\pm$0.05)	0.56 ($\pm$0.16)	0.69 ($\pm$0.02)
	Modularity	**0.67** ($\pm$0.05)	0.52 ($\pm$0.14)	0.63 ($\pm$0.01)
	No. of cluster accuracy	**0.64** ($\pm$0.17)	0.56 ($\pm$0.08)	0.55 ($\pm$0.07)

Table 4 Comparison of proposed algorithm in synthetic dataset with AFFECT k-means and AFFECT spectral in five measures of (a) Rand Index, (b) NMI, (c) F, (d) modularity and (e) accuracy of the no. of clusters

Network event	Measure	Granular-ARTISON	AFFECT k-means	AFFECT spectral
Merging and splitting	Rand Index	**0.95** ($\pm$0.02)	0.65 ($\pm$0.21)	0.75 ($\pm$0.20)
	NMI	**0.78** ($\pm$0.11)	0.49 ($\pm$0.16)	0.61 ($\pm$0.20)
	F-measure	**0.78** ($\pm$0.11)	0.45 ($\pm$0.19)	0.56 ($\pm$0.25)
	Modularity	**0.60** ($\pm$0.03)	0.40 ($\pm$0.20)	0.56 ($\pm$0.17)
	No. of clusters accuracy	**0.82** ($\pm$0.16)	0.61 ($\pm$0.03)	0.63 ($\pm$0.01)

results of this experiment. Each experiment is executed for 50 times and average values are reported.

The high difference of the number of clusters found compared to ground truth value in both Link-Clustering and OCG value are verified in these experiments. This problem in is reported in different papers (Ding et al. 2016) and demands more attention. Obviously, Granular-ARTISON shows the best performance in all scenarios of this experiment (Table 7). The average value of the number of clusters is reported in Table 7.

4.3 Real dataset evaluation

In this section, we evaluate the performance of Granular-ARTISON on several well-known benchmarks used heavily for evaluation of community detection algorithms. Networks studied are Zachary Karate club (Zachary 1977), Doubtful Sound Dolphins (Lusseau et al. 2003), and the Risk Game network (Steinhaeuser and Chawla 2010). These benchmarks assess the performance of the algorithms in static settings. For the purpose of the evaluation in dynamic real networks, we use the weighted blog dataset (Lin et al. 2008) provided by NEC laboratory gathered during 15 months of monitoring. The dataset was recently used in numerous studies of dynamic community detection (Lin et al. 2008; Yang et al. 2011). For a comparison with the state-of-the-art algorithms in different category of community detection, two other algorithms, i.e., Louvain (Blondel et al. 2008) and SLPA (Xie et al. 2011)/LabelRankT (Xie et al. 2013), are used. Louvain is a modularity-based algorithm, and SLPA and LabelRankT are placed in the label propagation-based category. In SLPA,

the probability of observing a label is interpreted as the membership strength. This algorithm can determine the number of communities automatically and is applicable to both weighted and directed versions. All algorithms can deal with weighted networks. Further, LabelRankT, which is proposed by the same author of the SLPA algorithm, is specifically designed for dynamic settings and is added for comparison with the dynamic NEC dataset. Table 8 gives a summary of the statistics on the datasets used for the evaluation of our algorithm in this paper.

4.3.1 Zachary Karate club

The well-known Karate club network shows the friendship networks of Football College. The experiments on this network split the network perfectly into two partitions without any mismatch in different measures of Rand Index, F-measure and NMI. The modularity of Granular-ARTISON is also higher than other algorithms. Finally, the number of clusters is best assessed in Granular-ARTISON and AFFECT spectral in common (Table 9, Fig. 11).

4.3.2 Dolphins

The other social network studied to test the accuracy of the algorithm is a social network of frequent associations among 62 dolphins in a community in Doubtful Sound (Lusseau et al. 2003). In this network, dolphins are represented as vertices, and a link is attached between two nodes if the corresponding dolphins are observed together more often than expected by chance over a period of 7 years from 1994 to 2001. The
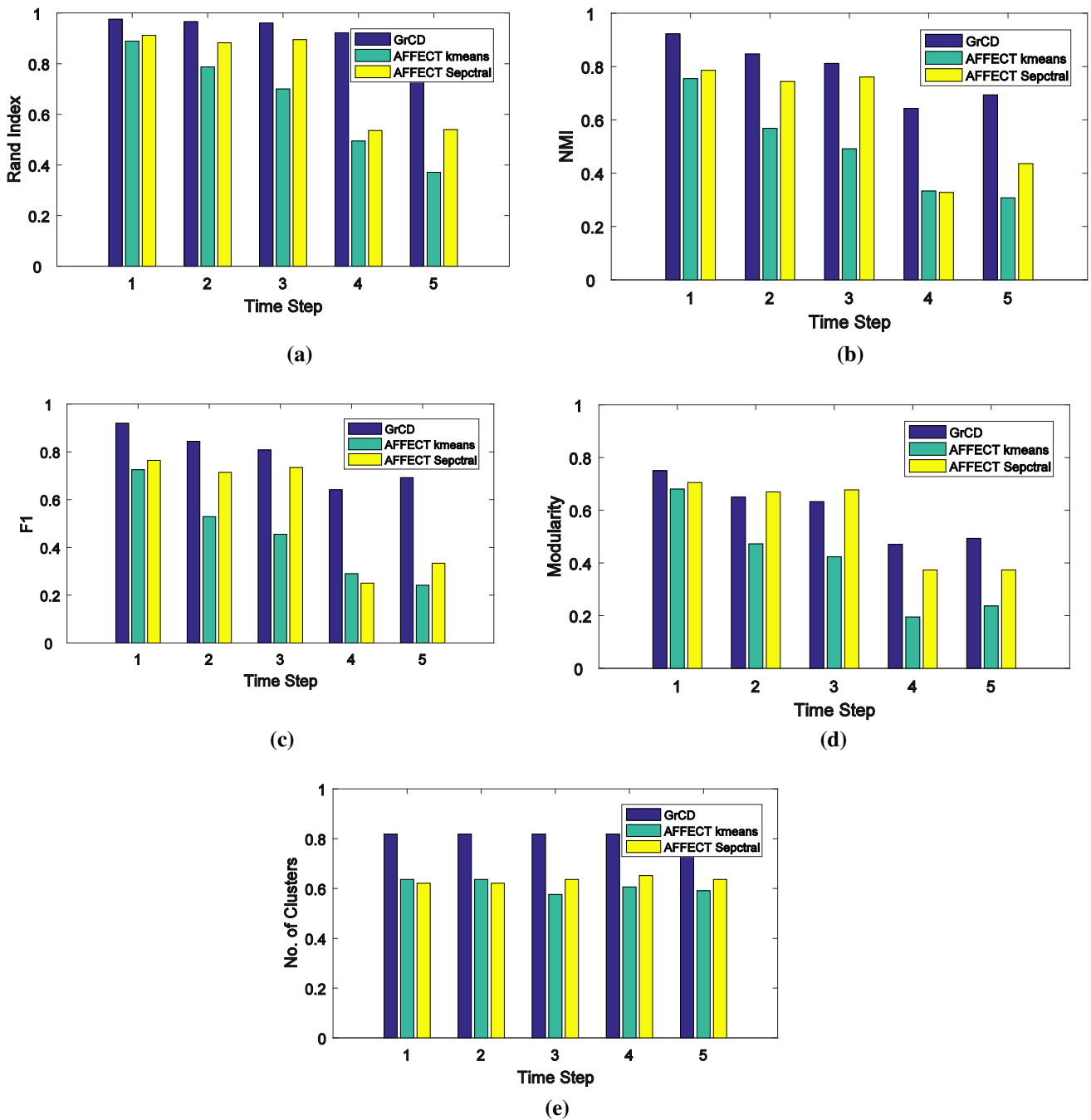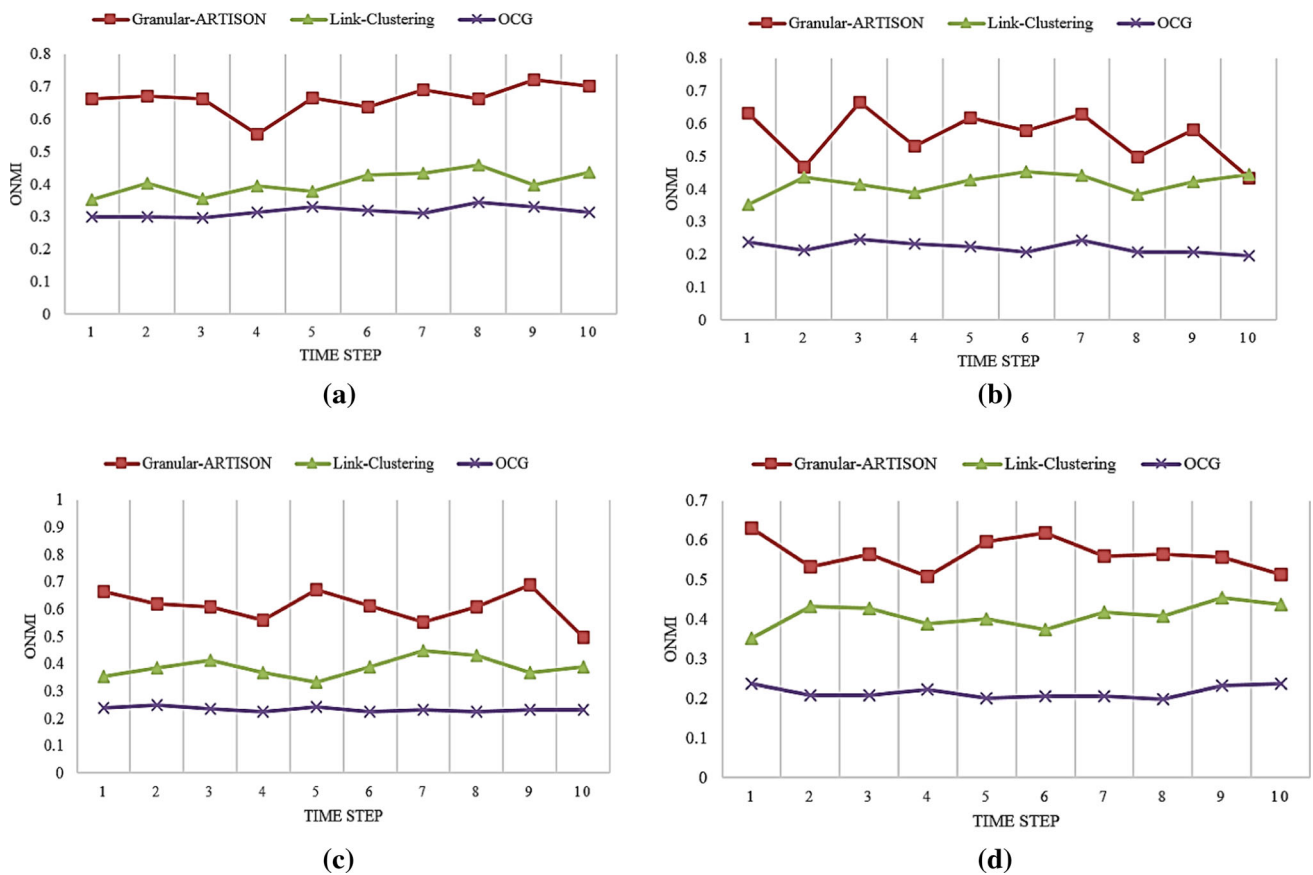
Fig. 8 Performance comparison of Granular-ARTISON in the *merging and splitting* event of the synthetic data generator with AFFECT k-means and AFFECT spectral in five measures of **a** Rand Index, **b** NMI, **c** F, **d** modularity and **e** accuracy of the no. of clusters

		Birth and death	Expansion	Hiding	Merge and split
Table 5 T-test results of average F-measure between Granular-ARTISON and the other evolutionary algorithm on four dynamic events	Granular-ARTISON and AFFECT k-means	0.018344	8.11461E−05	0.00418	0.00128
	Granular-ARTISON and AFFECT spectral	0.000879	0.084382	0.00191	0.02585

Fig. 9 Performance evaluation of Granular-ARTISON in overlapping mode in the four scenarios of **a** *birth and death*, **b** *expansion and contraction*, **c** *hiding* and **d** *merge and split event* versus Link-Clustering and OCG algorithms in overlapping-NMI measure

Table 6 Average value and standard deviation value of Granular-ARTISON in overlapping mode in the four scenarios of (a) *birth and death*, (b) *expansion and contraction*, (c) *hiding* and (d) *merge and split events* versus Link-Clustering and OCG algorithms in overlapping-NMI measure

	Birth and death	Expansion and contraction	Hiding	Merge and split
Granular-ARTISON	**0.66 (± 0.04)**	**0.56 (± 0.07)**	**0.60 (± 0.03)**	**0.56 (± 0.03)**
Link-clustering	0.40 (± 0.03)	0.51 (± 0.04)	0.38 (± 0.03)	0.40 (± 0.03)
OCG	0.31 (± 0.01)	0.22 (± 0.01)	0.23 (± 0.01)	0.21 (± 0.01)

groups of dolphins are mainly divided into the male ones and female ones.

Our result is shown to be completely the same as the ground truth. The two communities are marked by purple and blue, respectively (shown in Fig. 12, Table 10).

4.3.3 Risk game network

The risk game network is recently used for assessing community detection accuracy of the algorithms (Gupta et al. 2016). It is a popular strategy game played on a board depicting a political map of the Earth, divided into forty-two territories which are grouped into six continents (Table 11).

4.3.4 NEC blog dataset

The temporal dataset is gathered over a 15-month period by an in-house blog crawler of NEC laboratory (Lin et al. 2008). It contains information on 407 blogs which contribute to 148,681 links to each other representing the interaction among individuals, e.g., hyperlinks in blogs. These interactions provide dynamic communities over time. Ground truth information is available for this dataset. The results of the experiment are illustrated in Fig. 13.

The results of this experiment on this large real network show a very promising result. In addition to being superior in almost all the measures, Granular-ARTISON preserves the highest difference in the values of the measures compared

(a)

(b)

(c)

(d)
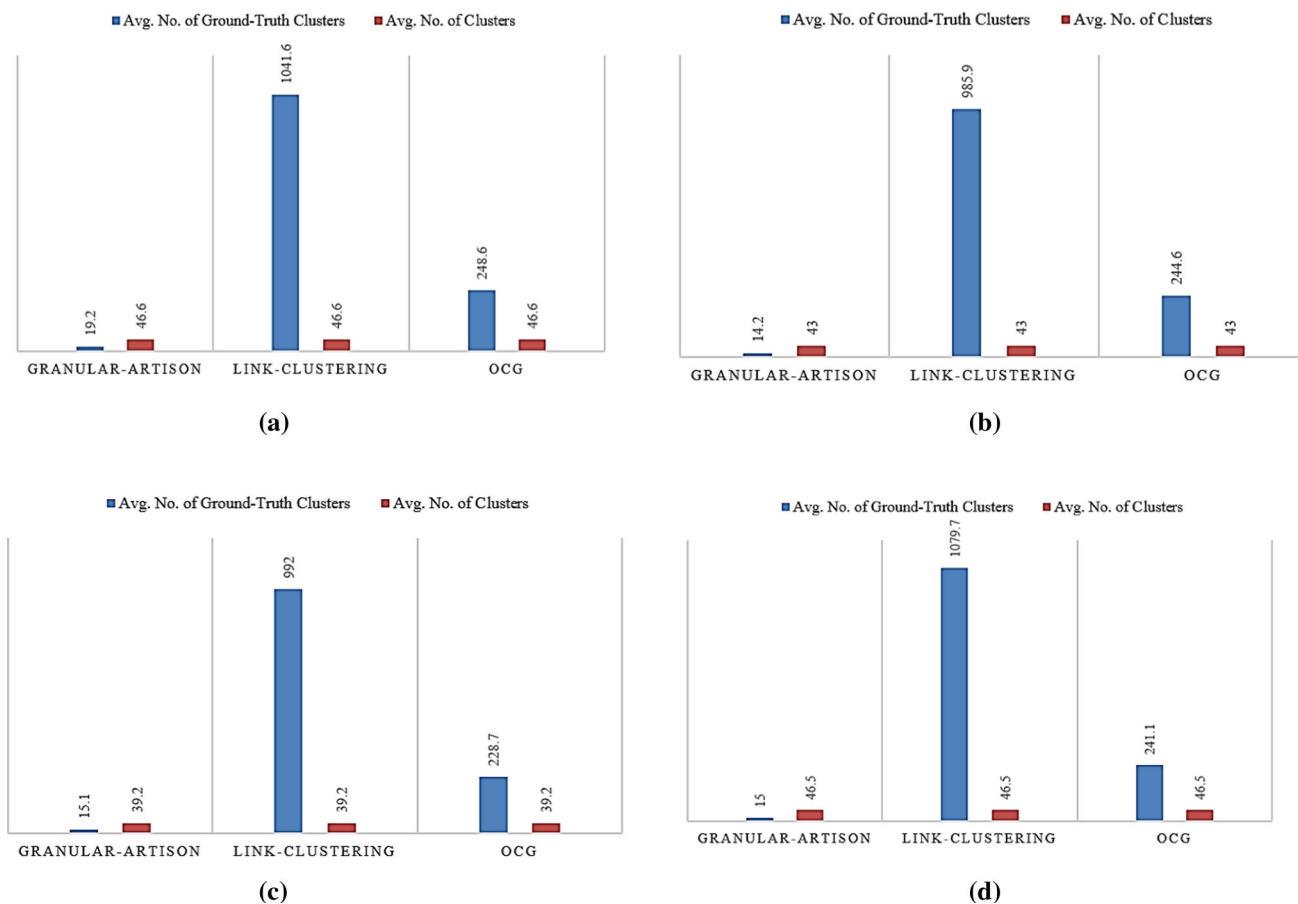
Fig. 10 Performance evaluation of Granular-ARTISON in overlapping mode in the four scenarios of **a** *birth and death*, **b** *expansion and contraction*, **c** *hiding* and **d** *merge and split event* versus Link-Clustering and OCG algorithms in the accuracy of the average number of clusters

Table 7 Average value and standard deviation value of Granular-ARTISON in overlapping mode in the four scenarios of (a) *birth and death*, (b) *expansion and contraction*, (c) *hiding* and (d) *merge and split events* versus Link-Clustering and OCG algorithms in the accuracy of the average number of clusters

	Birth and death	Expansion and contraction	Hiding	Merge and split
Ground truth	46.6 ($\pm$2.53)	43.0 ($\pm$2.98)	39.2 ($\pm$1.08)	46.5 ($\pm$2.98)
Granular-ARTISON	**19.2 ($\pm$2.69)**	**14.2 ($\pm$3.23)**	**15.1 ($\pm$3.50)**	**15.0 ($\pm$2.63)**
Link-clustering	1041.6 ($\pm$148.93)	985.9 ($\pm$115.16)	999.0 ($\pm$132.27)	1079.7 ($\pm$157.95)
OCG	248.6 ($\pm$13.25)	244.65 ($\pm$17.95)	228.7 ($\pm$2.51)	241.1 ($\pm$13.25)

Table 8 Statistics of the datasets used for experiments

| Network dataset	# Nodes	# Edges	Average degree
Karate	34	78	4.5
Risk	42	165	5.5
Dolphins	62	165	5.1
NEC Blog	407	148,681	8.3

with other algorithms. Louvain is the only algorithm which has a higher value of modularity, which is expected from this algorithm. The comparison among the four other algorithms inferior to Granular-ARTISON is as follows: interestingly, AFFECT k-means shows a better result than AFFECT spectral. The results indicate that AFFECT k-means in most
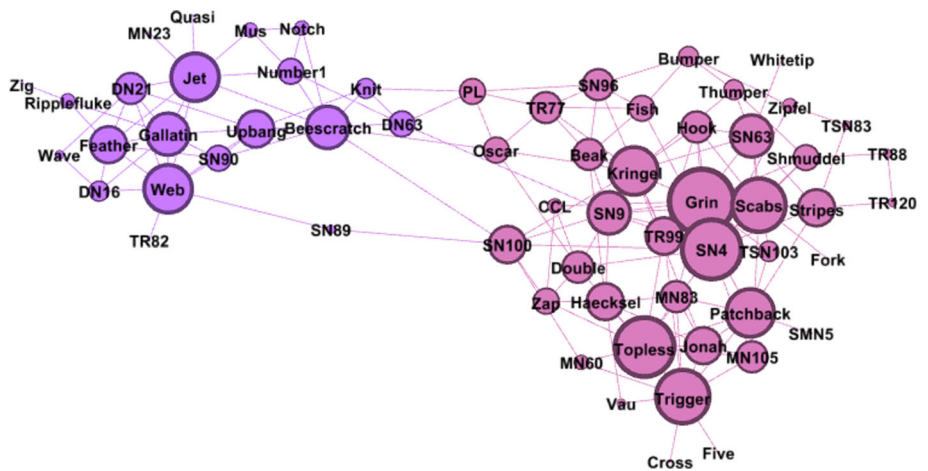
Table 9 Comparison of Granular-ARTISON in *Zachary club football* dataset with AFFECT k-means, AFFECT spectral, Louvain and SLPA in mean value of five measures: Rand Index, NMI, F, modularity and the accuracy of the assessed number of communities

	Granular-ARTISON	AFFECT k-means	AFFECT Spectral	Louvain	SLPA
Rand Index	**1**	0.84	0.94	0.66	0.68
NMI	**1**	0.65	0.84	0.55	0.45
F-measure	**1**	0.69	0.84	0.49	0.44
Modularity	**0.37**	0.26	0.36	0.32	0.22
No. of clusters accuracy	**1**	0.2	1	0.25	0.75

Fig. 11 Community structure of Zachary club football network discovered by Granular-ARTISON in two cases of **a** disjoint clusters and **b** overlapping clusters (purple items are common in overlapping regions of two communities) (color figure online)

(a)   **(b)**
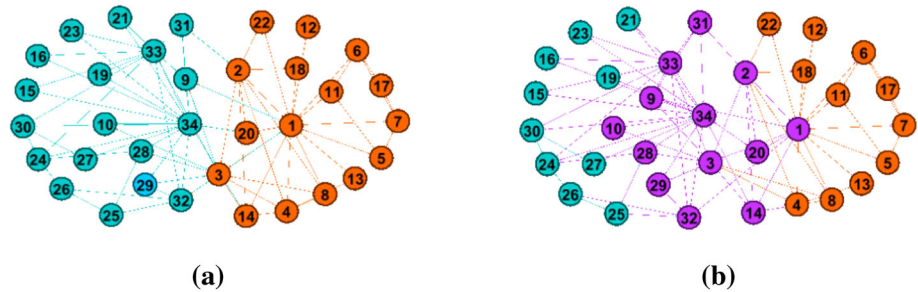
Fig. 12 Community structure of Dolphin network discovered by Granular-ARTISON

Table 10 Comparison of Granular-ARTISON in *dolphins' network* dataset with AFFECT k-means, AFFECT spectral, Louvain and SLPA in mean value of five measures: Rand Index, NMI, F, modularity and the accuracy of the assessed number of communities

	Granular-ARTISON	AFFECT k-means	AFFECT Spectral	Louvain	SLPA
Rand Index	0.74	0.69	0.74	0.73	**0.94**
NMI	**0.65**	0.49	0.71	0.56	0.55
F-measure	**0.64**	0.44	0.66	0.56	0.55
Modularity	0.38	0.34	0.38	0.24	**0.40**
No. of clusters accuracy	**0.88**	0.75	0.75	0.75	0.75

all others in this dataset. For NMI and F-measures, Granular-ARTISON always outperforms the dynamic k-means-based algorithms. In particular, in the best-tuned measure for accuracy, which penalizes both false negative and false positive results; i.e., F-measure, Granular-ARTISON shows its superiority with regard to other algorithms by 40% in the worst case and 60% in the best case. The average and slight variations in the results in this set of experiments are reported in Table 12.

Finally, we experiment the effect of using incremental community discovery in the proposed algorithm. As illustrated in Fig. 14 and Table 13, when the community structure in each time step is initialized using the results derived from the previous time step, the algorithm shows higher performance and the changes of the results in the consequent time steps are smoother (on average 16% improvement is obtained).

measures is ranked second, and Louvain is in third place. AFFECT spectral in NMI, F and modularity index is in fourth place. The LabelRankT algorithm works worse than

Table 11 Comparison of Granular-ARTISON in *risk map* dataset with AFFECT k-means, AFFECT spectral, Louvain and SLPA in mean value of five measures: Rand Index, NMI, F, modularity and the accuracy of the assessed number of communities

	Granular-ARTISON	AFFECT k-means	AFFECT Spectral	Louvain	SLPA
Rand Index	**0.87**	0.70	0.63	0.85	0.73
NMI	**0.71**	0.52	0.71	0.56	0.55
F1 measure	0.75	0.53	0.49	0.72	**0.90**
Modularity	0.60	0.43	0.39	0.43	**0.62**
No. of clusters accuracy	0.92	0.75	0.67	0.58	**1**

5 Conclusion and future works

We have proposed a dynamic community detection algorithm called Granular-ARTISON based on granulation concepts and our previous human-inspired community detection algorithm. Granular-ARTISON works using local information and functions in dynamic/static and overlapping/non-overlapping contexts. These features are particularly important in real social networks and very few algorithms deal with all of them simultaneously. Like the incremental mining algorithms, Granular-ARTISON exploits the previously discovered community prototypes for new community discovery
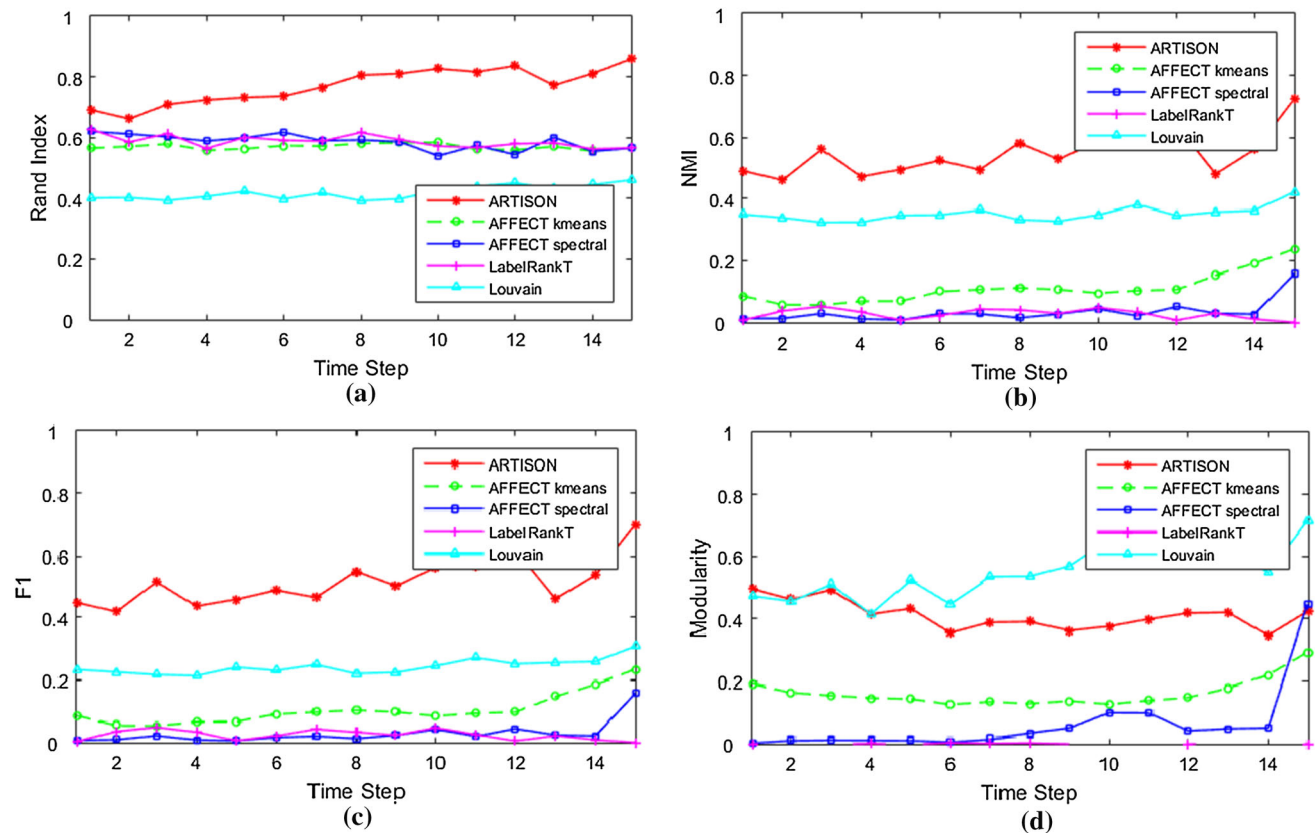

Fig. 13 Performance comparison of Granular-ARTISON in *NEC blog* dataset with AFFECT k-means, AFFECT spectral, Louvain and LabelRankT in four measures of **a** Rand Index, **b** NMI, **c** F-measure and **d** modularity

Table 12 Comparison of the proposed algorithm in NEC dataset with AFFECT K-MEANS, AFFECT spectral, Louvain and fast modularity in mean value of five measures: Rand Index, NMI, Precision and F and Modularity

	Granular-ARTISON	AFFECT k-means	AFFECT spectral	Louvain	LabelRankT
Rand Index	**0.87 (±0.04)**	0.57 (±0.01)	**0.59 (±0.03)**	**0.42 (±0.02)**	**0.59 (±0.02)**
NMI	**0.62 (±0.08)**	**0.62 (±0.14)**	**0.12 (±0.05)**	**0.35 (±0.02)**	**0.03 (±0.03)**
F-measure	**0.61 (±0.08)**	**0.11 (±0.05)**	**0.03 (±0.04)**	**0.24 (±0.02)**	**0.02 (±0.02)**
Modularity	**0.29 (±0.05)**	**0.17 (±0.05)**	**0.07 (±0.05)**	**0.56 (±0.10)**	0.0 (±0.0)

Fig. 14 Performance evaluation of Granular-ARTISON in F1 measure where two scenarios is considered: (a) no initialization of the previous state of the network is considered (blue line) and (b) the algorithm is initialized in each time step by the network results obtained in the previous time step (red line) (color figure online)
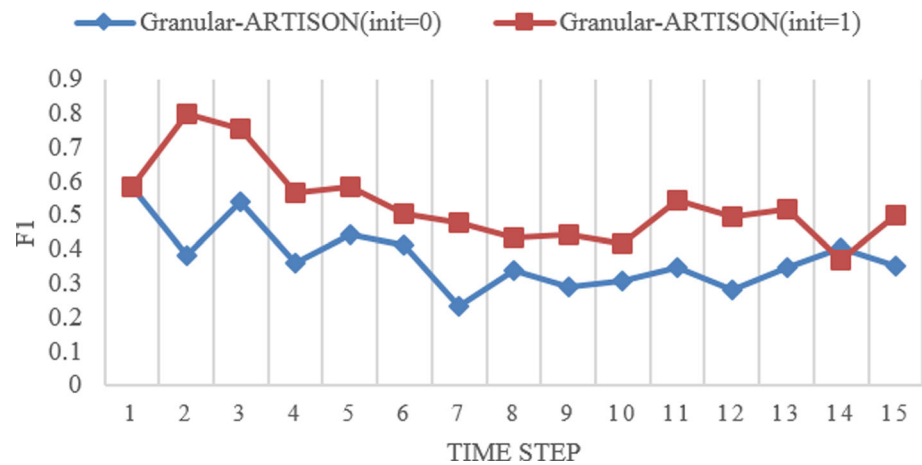
Table 13 Average value of F-measure in two set of experiments evaluating the forget factor presence and incremental initialization influence in Granular-ARTISON algorithm

	Incremental initialization	No incremental initialization
Average F-measure	**0.53 (± 0.09)**	**0.37 (± 0.03)**

to preserve temporal smoothness. However, the granulation mechanism in both lower and higher levels of the algorithm makes our representative-based algorithms unique in different aspects. Specifically, the algorithm can deal with both low and abrupt changes in the network while being able to determine the correct number of communities automatically. We used extensive evaluation to show the effectiveness of our Granular-ARTISON model against state-of-the-art algorithms on both real and synthetic datasets. In almost all cases, the results of the experiments confirmed the superiority of Granular-ARTISON in different measures. Meanwhile, several improvements are being researched for this algorithm. Designing proper meet operations among macrogranules to split larger communities into smaller ones due to shrinkage and deriving quantified membership degree for communities is under investigation. Finally, since the algorithm works totally by local information, the experiments for running the algorithm in the streaming mode are under study.

Compliance with ethical standards

Conflict of interest No conflict of interest is associated with this article.

References

Ahn Y-Y, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. Nature 466(7307):761–764

Amelio A, Pizzuti C (2014) Overlapping community discovery methods: a survey. In: Gündüz-Öğüdücü Ş, Etaner-Uyar A (eds) Social networks: analysis and case studies. Lecture notes in social networks. Springer, Vienna, pp 105–125

Becker E, Robisson B, Chapple CE, Guénoche A, Brun C (2011) Multifunctional proteins revealed by overlapping clustering in protein interaction network. Bioinformatics 28(1):84–90

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 10:P10008

Breve F, Zhao L (2013) Fuzzy community structure detection by particle competition and cooperation. Soft Comput 17(4):659–673

Cazabet R, Amblard F (2014) Dynamic community detection. In: Alhajj R, Rokne J (eds) Encyclopedia of social network analysis and mining. Springer, New York, NY, pp 404–414. https://doi.org/10.1007/978-1-4614-6170-8_383

Chakrabarti D, Kumar R, Tomkins A (2006) Evolutionary clustering. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 554–560

Chi Y, Song X, Zhou D, Hino K, Tseng BL (2007) Evolutionary spectral clustering by incorporating temporal smoothness. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 153–162

Dillen NB, Chakraborty A (2016) Modularity-based community detection in fuzzy granular social networks. In: Proceedings of the international congress on information and communication technology. Springer, pp 577–585

Ding Z, Zhang X, Sun D, Luo B (2016) Overlapping community detection based on network decomposition. Sci Rep 6:24115

Fahmi A, Abdullah S, Amin F, Ali A (2018) Weighted average rating (War) method for solving group decision making problem using triangular cubic fuzzy hybrid aggregation (Tcfha). Punjab Univ J Math 50(1):23–34

Folino F, Pizzuti C (2014) An evolutionary multiobjective approach for community discovery in dynamic networks. IEEE Trans Knowl Data Eng 26(8):1838–1852

Görke R, Maillard P, Schumm A, Staudt C, Wagner D (2013) Dynamic graph clustering combining modularity and smoothness. J Exp Algorithmics (JEA) 18(1):1–5

Greene D, Doyle D, Cunningham P (2010) Tracking the evolution of communities in dynamic social networks. In: 2010 international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 176–183

Grossberg S (2013) Adaptive resonance theory: how a brain learns to consciously attend, learn, and recognize a changing world. Neural Netw 37:1–47

Gupta S, Kumar P, Bhasker B (2016) A rough connectedness algorithm for mining communities in complex networks. In: International conference on big data analytics and knowledge discovery. Springer, pp 34–48

Hartmann T, Kappes A, Wagner D (2016) Clustering evolving networks. In: Kliemann L, Sanders P (eds) Algorithm engineering. Springer, Cham, pp 280–329

Kundu S, Pal SK (2015) FGSN: fuzzy granular social networks-model and applications. Inf Sci 314:100–117

Lin Y-R, Chi Y, Zhu S, Sundaram H, Tseng BL (2008) Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. In: Proceedings of the 17th international conference on World Wide Web. ACM, pp 685–694

Lin Y-R, Chi Y, Zhu S, Sundaram H, Tseng BL (2009) Analyzing communities and their evolutions in dynamic social networks. ACM Trans Knowl Discov Data (TKDD) 3(2):8

Lingras P, West C (2004) Interval set clustering of web users with rough k-means. J Intell Inf Syst 23(1):5–16

Liu H, Liu C, C-a Wu (2015) A framework of granular computing clustering algorithms. Int J Hybrid Inf Technol 8(12):225–230

Lusseau D, Schneider K, Boisseau OJ, Haase P, Slooten E, Dawson SM (2003) The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Behav Ecol Sociobiol 54(4):396–405

McDaid AF, Greene D, Hurley N (2011) Normalized mutual information to evaluate overlapping community finding algorithms. arXiv preprint arXiv:11102515

Nguyen NP, Dinh TN, Tokala S, Thai MT (2011) Overlapping communities in dynamic networks: their detection and mobile applications. In: Proceedings of the 17th annual international conference on Mobile computing and networking. ACM, pp 85–96

Oner SC, Oztaysi B (2018) An interval type 2 hesitant fuzzy MCDM approach and a fuzzy c means clustering for retailer clustering. Soft Comput 22:4971–4987

Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818

Peters G, Weber R (2009) Intelligent cluster algorithms for changing data structures. Int J Intell Defence Support Syst 2(2):105–119

Peters G, Weber R (2016) DCC: a framework for dynamic granular clustering. Granul Comput 1(1):1–11

Peters G, Crespo F, Lingras P, Weber R (2013) Soft clustering–fuzzy and rough approaches and their extensions and derivatives. Int J Approx Reason 54(2):307–322

Plantié M, Crampes M (2013) Survey on social community detection. In: Ramzan N, van Zwol R, Lee J-S, Clüver K, Hua X-S (eds) Social media retrieval. Springer, London, pp 65–85

Rosvall M, Bergstrom CT (2010) Mapping change in large networks. PLoS ONE 5(1):e8694

Sawyer RK (2005) Social emergence: societies as complex systems. Cambridge University Press, Cambridge

Steinhaeuser K, Chawla NV (2008) Community detection in a large real-world social network. In: Liu H, Salerno JJ, Young MJ (eds) Social computing, behavioral modeling, and prediction. Springer, Boston, pp 168–175

Steinhaeuser K, Chawla NV (2010) Identifying and evaluating community structure in complex networks. Pattern Recognit Lett 31(5):413–421

Takaffoli M, Fagnan J, Sangi F, Zaïane OR (2011) Tracking changes in dynamic information networks. In: 2011 international conference on computational aspects of social networks (CASoN) IEEE, pp 94–101

Tang L, Liu H, Zhang J (2012) Identifying evolving groups in dynamic multimode networks. IEEE Trans Knowl Data Eng 24(1):72–85

Wang W, Liu D, Liu X, Pan L (2013) Fuzzy overlapping community detection based on local random walk and multidimensional scaling. Physica A 392(24):6578–6586

Whang JJ, Gleich DF, Dhillon IS (2016) Overlapping community detection using neighborhood-inflated seed expansion. IEEE Trans Knowl Data Eng 28(5):1272–1284

Xie J, Szymanski BK, Liu X (2011) Slpa: uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In: 2011 IEEE 11th international conference on data mining workshops. IEEE, pp 344–349

Xie J, Chen M, Szymanski BK (2013) LabelrankT: Incremental community detection in dynamic networks via label propagation. Proceedings of the workshop on dynamic networks management and mining. ACM

Xie J, Kelley S, Szymanski BK (2013b) Overlapping community detection in networks: the state of the art and comparative study. ACM Comput Surv 45(4):43

Xu KS, Kliger M, Hero Iii AO (2014) Adaptive evolutionary clustering. Data Min Knowl Disc 28(2):304–336

Yager RR, Filev D (1998) Operations for granular computing: mixing words and numbers. In: The 1998 IEEE international conference on fuzzy systems proceedings. IEEE World Congress on Computational Intelligence. IEEE, pp 123–128

Yang T, Chi Y, Zhu S, Gong Y, Jin R (2011) Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. Mach Learn 82(2):157–189

Zachary WW (1977) An information flow model for conflict and fission in small groups. J Anthropol Res 33:452–473

Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90(2):111–127