ISSN: 0067-2904

Community Tracking in Time Evolving Networks: An Evolutionary Multi-objective Approach

Haidar S. Khoder, Bara A. Attea*
Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq.

Abstract

In real world, almost all networks evolve over time. For example, in networks of friendships and acquaintances, people continually create and delete friendship relationship connections over time, thereby add and draw friends, and some people become part of new social networks or leave their networks, changing the nodes in the network. Recently, tracking communities encountering topological shifting drawn significant attentions and many successive algorithms have been proposed to model the problem. In general, evolutionary clustering can be defined as clustering data over time wherein two concepts: snapshot quality and temporal smoothness should be considered. Snapshot quality means that the clusters should be as precise as possible during the current time step. Temporal smoothness, on the other hand, means that the clusters should not changed dramatically between successive time steps. In this paper, a multi-objective optimization model, based on internal community density as snapshot metric, is proposed and compared with the state-of-the-art modularity based model. Both models are then used to solve the community tracking problem in dynamic social network. The problem, in both models, is stated as a multi-objective optimization problem and the decomposition based multi-objective evolutionary algorithm is used to solve the problem. Experimental results reveals that the proposed model significantly outperforms the already existing model in the ability of tracking more shifted communities.

Keywords: Evolutionary clustering; evolutionary network analysis; dynamic social networks; graph partitioning; social network analysis.

تعقب المجتمع في الشبكات الاجتماعية المتغيره مع مرور الزمن: طريقة تطورية متعددة الأهداف

حيدر سفير خضير، براء علي عطية*
قسم الحاسوب ، كلية العلوم ، جامعة بغداد ، بغداد ، العراق

الخلاصة

معظم الشبكات في العالم الحقيقي تتطور وتتغير مع مرور الزمن، على سبيل المثال الشبكات التي يتم من خلالها التعارف، طبيعة هذه الشبكات تتغير مرارا مع مرور الزمن ، و الصداقات، عن طريق انشاء علاقات وروابط جديدة او العكس، وبذلك أصبح بعض الاشخاص جزءا من شبكات اجتماعية جديدة أو ترك شبكاتهم التي كانو يرتبطون بها مسبقا. مؤخرا هنالك العديد من الخوارزميات المتلاحقه التي اقترحت لتعالج مشكله تعقب واكتشاف المجتمعات التي تتميز ب التحول الطوبوغرافي. عموما، الخوارزمية التطويرية للكشف عن المجتمعات يمكن تعريفها بانها تجميع البيانات مع مرور الزمن. هنالك مصطلحين مهمين في الخوازميه التطوريه للكشف عن المجتمعات هما : جودة الانتقاء و السلاسة الزمنية . جودة الا نتقاء تعني ايجاد

*Email: baraali@scbaghdad.edu.iq

المجتمعات يجب أن يكون دقيقا قدر الإمكان أثناء الفتره الزمنية الحالية. اما السلاسة الزمنية تعرّف بان المجتمعات يجب ان لاتتغير بشكل كبير بين الزمن الحالي والزمن السابق له . وبصيغه اخرى يمكن ان تعرف بأن التجمعات يجب ان لاتتغير بشكل كبير بين الخطوات الزمنيه المتتاليه . في هذا البحث. تم أقتراح نموذج متعدد الاهداف، بالاعتماد على الكثافة الداخليه للمجتمعات كأنتقاء قياسي، حيث اقترح وقورن مع نموذج اخر تمت صياغته في الأدبيات يعرف بنوذج التجزئه. كلاهما طبقا عمليا لحل مشكلة اكتشاف المجتمعات داخل الشبكات الديناميكيه المعقده المتغيره مع الزمن . في كلا النموذجين، المشكله محدده على اساس مشكله متعددة الاهداف بالاعتماد على الخوارزميه التطوريه متعددة الاهداف لحل هذه المشكله. النتائج التجريبية تكشف أن النموذج المقترح يتفوق بشكل كبير على النموذج القائم بالفعل في القدرة على اكتشاف المجتمعات المتغيره مع الزمن.

1. Introduction

Over the past few years, the application of dynamic social network has steadily grown interests. It has become an important part in many disciplines that involve dynamic systems. Examples include, but not limited to, social-communication networks, biological networks, World Wide Web, Face-book and Twitter, etc. Evolutionary clustering of a dynamic network aims at mining *evolving pattern* of membership of individuals (denoted as module, community, or simply, cluster) that has dense (intra-connection) between them and sparse (inter-connection) with other patterns. From this perspective, evolutionary clustering can emerge to find evolving communities over time. Evolutionary clustering having temporal phenomenon is first introduced by Chakrabarti et al. [1]. They stated that evolutionary clustering can be modeled while simultaneously optimizing two criteria: snapshot quality (where a cluster should reflect accurate data membership at the current time) and, temporal smoothness (where a cluster could shift, but smoothly, over time).

Mathematically, a network is modeled as graph of pairwise edges between its nodes. Consider a social network $\mathcal{N}$ of n individuals being modeled by $G = (V, E)$. Let $\mathcal{C} = \{C_1, \dots, C_K\}$ be a candidate partitioning of $\mathcal{N}$. Let n_i and $m_i = \sum_{v,w \in C_i}(v, w)$ be the degree and volume of C_i, respectively. Moreover, let $in_i(v) = \sum_{w \in C_i}(v, w)$ and $out_i(v) = \sum_{\mathrm{w} \notin \mathrm{C_i}}(\mathrm{v}, \mathrm{w})$ be, respectively, the number of intra-connections and inter-connections of node v which belongs to cluster C_i (i.e. $|l_v| = in_i(v) + out_i(v)$). Figure-1 captures the evolution of a Kim and Han [2] social network, consisting of 128 nodes. The figure captures the network at 10 different time steps.

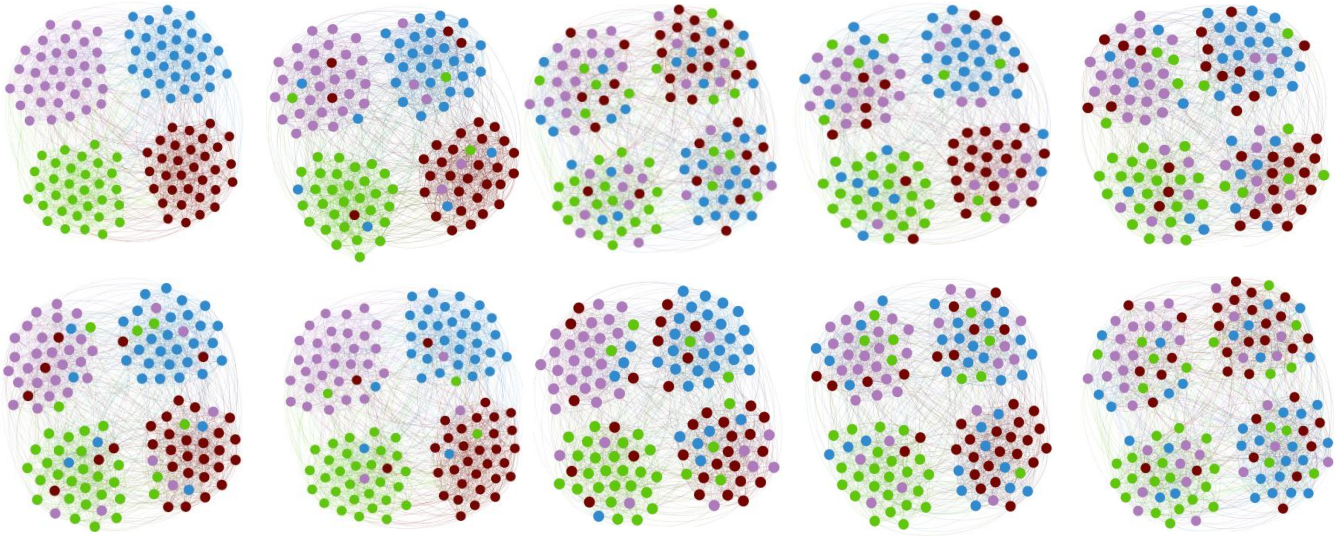

Figure 1- Evolution of Kim and Hans' 128-node network with 4 communities at ten time steps (from left to right and from top to bottom).

Chakrabarti et al. [1] proposed a two-fold evolutionary clustering framework to maintain both snapshot cost (SC) and temporal cost (TC) by simultaneously optimizing two conflicting criteria: the accuracy of clustering at the current time step and the gradual drift from the most recent clustering history. Maximizing both SC and TC reflects, as a result, the mild transition of nodes through successive time steps. Inspired by the work of Chakrabarti et al., several heuristic and meta-heuristic evolutionary clustering approaches have been proposed in the literature with paramount performance for the multi-objective evolutionary algorithms ($MOEAs$). For example, a multi-objective evolutionary based evolutionary clustering algorithm is recently proposed by Folino and Pizzuti [3] and [4], to outperform most state-of-the-art methods. Snapshot quality is maintained by adopting the most common measures used in the literature to capture the hidden structure of community. Snapshot cost, SC, is measured using $modularity$ [5], $community\ score$ [4], $conductance$ [6], or $normalized\ cut$ [7]. They demonstrated that $MOEAs$ deserve the credit for providing additional improvement on the accuracy of evolutionary clustering over other state-of-the-art methodologies. Moreover, they proved that modularity based $MOEA$ (Q) has the best performaance over other $MOEA$'s models.

$Modularity$, Q [5], awards a partition solution $\mathcal{C} = \{C_1, \dots, C_K\}$ according to the fraction of intra-connections inside $\{C_1, \dots, C_K\}$ (Eq. 1). In Eq. 1, two contradictory objectives are handled. The first term in Eq. 1 biases towards a solution $\mathcal{C}$ with a densely intra-connected modules. On the other hand, the second term expresses that the expected value of the same edge density in $\mathcal{C}$ with the same community structure $\{C_1, \dots, C_K\}$ but fall at random between the vertices should be small. Q will approach its minimum at 0 if the number of within-community edges is no better than random. On the other hand, values approaching $Q = 1$, which is the maximum, indicate strong community structure.

$$max\ Q(\mathcal{C}_t) = \sum_{i=1}^{K} \left[\frac{m_i}{|L|} - \left(\frac{\sum_{v \in C_i} |l_v|}{2|L|}\right)^2\right] \quad (1)$$

On the other hand, temporal cost, TC, (using $Normalized\ Mutual\ Information - NMI$ [6]) measures how similar the community structure of a time step t to that at time $t-1$. NMI between two partitions $\mathcal{C}^t$ and $\mathcal{C}^{t-1}$ of a network $\mathcal{N}$ of n nodes at times t and $t-1$, respectively, is the normalization of the mutual information (MI) score between $\mathcal{C}^t$ and $\mathcal{C}^{t-1}$ being scaled between 0 (no mutual information) and 1.0 (perfect correlation). Consider the confusion matrix $c = [c_{ij}]$, $i = 1, \dots, K_{\mathcal{C}^t}$ and $j = 1, \dots, K_{\mathcal{C}^{t-1}}$, where c_{ij} be the number of nodes of community i of $\mathcal{C}^t$ that are also in community j of $\mathcal{C}^{t-1}$. Then,

$$NMI(\mathcal{C}^t, \mathcal{C}^{t-1}) = \frac{-2\sum_{i=1}^{K_{\mathcal{C}^t}} \sum_{j=1}^{K_{\mathcal{C}^{t-1}}} c_{ij} log(c_{ij} * n / c_i c_j)}{\sum_{i=1}^{K_{\mathcal{C}^t}} c_i log(c_i/n) + \sum_{j=1}^{K_{\mathcal{C}^{t-1}}} c_j log(c_j/n)} \quad (2)$$

where c_i and c_j are the sum of elements of community i in $\mathcal{C}^t$ and community j in $\mathcal{C}^{t-1}$, respectively. Note that when $t = 1$, temporal cost, TC, will be neglected from the whole formula and the problem will be stated as single objective community detection problem.

In terms of snapshot cost function (SC), however, some of other well known community detection models are still uninvestigated in the literature of $MOEA$ based evolutionary clustering algorithms. These includes $Internal\ Density$ (ID) and $Expansion$ (EX) [7], and $Cut\ Ratio$ (CR) [8]. The contribution of this paper is to formulate evolutionary clustering problem as a multi-objective optimization problem, based on $Internal\ Density$ (ID). Section 2 briefly review basic concepts of multi-objective evolutionary algorithms. This is followed by the proposed MOEA based formulation for the multi-objective evolutionary clustering problem. Section 4, then, evaluate the performance of the proposed evolutionary clustering model against Folino and Pizzuti's model [4]. The final section presents major conclusions and recommendation of this work.

2 Multi-objective evolutionary algorithms (MOEA)

Instead of single optimal or near-optimal solution, a set of non-dominated solutions can simultaneously be obtained to solve different real world problems as multi-objective optimization problems (MOPs). Consider a MOP of n decision variables $\mathbb{X} \in \mathbb{R}^n = [x_1, x_2, \dots, x_n]^T$, and k objective functions $\mathbb{F}(\mathbb{X}) = [\mathbb{f}_1(\mathbb{X}), \mathbb{f}_2(\mathbb{X}), \dots, \mathbb{f}_k(\mathbb{X})]^T$, i.e. $\mathbb{F}: \mathbb{R}^n \to \mathbb{R}^k$. $\mathbb{F}(\mathbb{X})$ is optimized (in terms of *domination*) towards finding Pareto-optimal set of solutions (also called Pareto set, PS), each of which is said to be a non-dominated or a non-inferior solution, denoted as $\mathbb{X}^* = [x_1^*, x_2^*, \dots, x_n^*]^T \mid \mathbb{X}^* \in \mathbb{X}$. To define domination, consider two vectors $\mathbb{U}$ and $\mathbb{V}$ from the solution space $\mathbb{R}^n = \Omega(\mathbb{X})$, i.e.

$\mathbb{U} \in \mathbb{X}$ and $\mathbb{V} \in \mathbb{X}$. Then, solution $\mathbb{U}$ is said to dominate $\mathbb{V}$ if and only if the following two conditions hold [9]:

1. Solution $\mathbb{U}$ is no worse than $\mathbb{V}$ in all objectives, or formally, $\forall i, 1 \leq i \leq k: \mathbb{f}_i(\mathbb{U}) \not\prec \mathbb{f}_i(\mathbb{V})$. For example in maximization, the word "no worse" means $\mathbb{f}_i(\mathbb{U}) \geq \mathbb{f}_i(\mathbb{V})$.

2. The solution $\mathbb{U}$ is *strictly* better than $\mathbb{V}$ in at least one objective, or formally, $\mathbb{f}_i(\mathbb{U}) \lhd \mathbb{f}_i(\mathbb{V})$ in at least one objective $\mathbb{f}_i$, $i \in \{1,2,\dots,k\}$. For maximization, the word "strictly better" means $\mathbb{f}_i(\mathbb{U}) > \mathbb{f}_i(\mathbb{V})$.

Hence, a non-dominated set can be defined as: among a set of solutions $\Omega(\mathbb{X})$, the non-dominated solutions set $\overline{\Omega}(\mathbb{X}) \subset \Omega(\mathbb{X})$ are subset of solutions which are not dominated by any other solution in $\Omega(\mathbb{X})$. Among the famous population based multi-objective evolutionary algorithms ($MOEAs$) being successfully applied to many real-world problems is the decomposition based multi-objective evolutionary algorithm ($MOEA/D$) of Zhang and Li [10]. Consider MOP with k objective functions:

$$min\ \mathbb{F}(\mathbb{X}) = [f_1(\mathbb{X}), f_2(\mathbb{X}), \dots, f_k(\mathbb{X})]^T \quad (3)$$
$$subject\ to\ \mathbb{X} \in \Omega$$

Also, consider a reference point $\mathbb{Z}^* = (z_1^*, \dots, z_k^*)$ to hold the best value obtained so far by $MOEA/D$ for each of the k objective functions, formally speaking: $\forall i \in \{1, \dots, k\}$

$$z_i^* = f_i(\mathbb{X}^* \in \overline{\Omega}(\mathbb{X})) : \Leftrightarrow \nexists\, \mathbb{X} \in\ \Omega(\mathbb{X}) |\ f_i(\mathbb{X}) > f_i(\mathbb{X}^*) \quad (4)$$
$$where\ \overline{\Omega} \subset \Omega$$

In $MOEA/D$, a population $\mathbb{P} = \{P_1, P_2, \dots, P_{PopSize}\}$ of parent solutions is used to represent $PopSize$ scalar optimization sub-problems. In other terms, MOP is decomposed by $MOEA/D$ into $PopSize$ sub-problems. Each individual $P_i, 1 \leq i \leq PopSize$ is associated with one weight vector λ_i of length k out of a set of $PopSize$ even spread weight vectors $\mathbb{W} = \{\lambda_1, \lambda_2, \dots, \lambda_{PopSize}\}$. Recall that there are k objective functions for the MOP, then each P_i has weight vector $\lambda_i = (\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,k}), s.t. \sum_{j=1}^{k} \lambda_{i,j} = 1$. Moreover, each individual $P_i, 1 \leq i \leq PopSize$ is evolved using information gathered only from its s neighbor solutions. Neighbor solutions to P_i, denoted by $\mathbb{g}_i = \{g_{i,1}, g_{i,2}, \dots, g_{i,s}\}$, are those with the closest distance (using Euclidean distance) weight vectors to λ_i. Formally, $\forall j \in \{1, \dots, s\}$

$$P_j \in \mathbb{g}_i : \Leftrightarrow \nexists\, \lambda_l \in \mathbb{W} | \sum_{x=1}^{k} (\lambda_{i,x} - \lambda_{j,x})^2 > \sum_{x=1}^{k} (\lambda_{i,x} - \lambda_{l,x})^2 \quad (5)$$

In $MOEA/D$, the problem of approaching close the Pareto Front, PF, of a general MOP defined in Eq. (3) can be decomposed into $PopSize$ scalar optimization sub-problems using Tchebycheff approach being formulated as: $\forall i \in \{1, \dots, PopSize\}\ \wedge\ \forall j \in \{1, \dots, k\}$

$$g_i^{te}(P_i | \lambda_i, z^*) = min_{1 \leq j \leq k} \{\lambda_{i,j} | f_j(P_i) - z_j^* |\} \quad (6)$$

In terms of minimization, $MOEA/D$ minimizes all these g_i^{te} scalar objective functions simultaneously in a single run. During evolution, $MOEA/D$ maintains an exterior population EP, for archiving non-dominated solutions found during the search. At each generation, $MOEA/D$ performs four main operations while generating new $PopSize$ solutions $\{P_1', P_2', \dots, P_{PopSize}'\}$:

- First, for each parent individual $P_i \in \mathbb{P}$, a new offspring solution P_i' is generated, using problem-specific genetic operators (e.g., crossover and mutation), from only its neighbors $\mathbb{g}_i$.
- Second, if necessary it updates the reference points $z^* = (z_1^*, \dots, z_k^*)$. $\forall j \in \{1, \dots, k\}$, if, in case of minimization, $z_j^* > f_j(P_i')$, then it sets $z_j^* = f_j(P_i')$.
- Third, it updates the neighbors of P_i: $\forall j \in \{1, \dots, s\}$, if $g_i^{te}(P_i' | \lambda_j, z^*) \leq g_j^{te}(g_{i,j} | \lambda_j, z^*)$, then it sets $g_{i,j} = P_i'$ and $\mathbb{F}(g_{i,j}) = \mathbb{F}(P_i')$.
- Finally, it updates EP by removing from it all solutions $\mathbb{Y}$ where $\mathbb{F}(P_i') \lhd \mathbb{F}(\mathbb{Y})$ and insert P_i' into EP if $\nexists \mathbb{Y} \in EP \rightarrow \mathbb{F}(\mathbb{Y}) \lhd \mathbb{F}(P_i')$.

Algorithm 1 outlines the general steps of $MOEA/D$.

Algorithm 1. **The general outline of MOEA/D**

Input:

- Multi-objective minimization problem $f(\mathbb{x}) = (f_1(\mathbb{x}), f_2(\mathbb{x}), \dots, f_m(\mathbb{x}))$
- Number of sub-problems to be evolved, i.e. population size, N
- Uniform spread of N weight vectors: $\lambda_1, \lambda_2, \dots, \lambda_N$ such that $\lambda_i = (\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,m})$
- Neighborhood size of each weight vector, T
- Maximum number of generations, max_t
- Probability of crossover, p_c
- Probability of mutation, p_m

Output: External archive of Pareto set of non-dominated solutions, EP.

Step 0 - Setup:

- $EP = \emptyset$
- $t = 0$

Step 1 – Initialization

- Uniformly, *generate* an initial population, $\mathbb{P}_0 = \{\mathbb{P}_1, \mathbb{P}_2, \dots, \mathbb{P}_N\}$.
- *Evaluate* fitness vector $f(\mathbb{P}_i) = (f_1(\mathbb{P}_i), f_2(\mathbb{P}_i), \dots, f_m(\mathbb{P}_i)), \forall i = 1,2, \dots, N$.
- *Initialize* ideal vector $z^* = (z_1^*, z_2^*, \dots, z_m^*)^T$ by a problem-specific method.
- *Compute* Euclidean distance between weight vectors $\lambda_1, \lambda_2 \dots, \lambda_N$and *assign* the T closest vectors $\lambda_i^1, \lambda_i^2, \dots, \lambda_i^T$to each λ_i. $\forall i = 1,2, \dots, N$, set $B(i) = \{i_1, i_2, \dots, i_T\}$.

Step 2 – Evolve cycle: For$i = 1,2, \dots, N$

- Randomly *select* two indices $k, l$from $B(i)$, and *generate* a new solution $\mathbb{Y}$ from $\mathbb{P}_k$ and $\mathbb{P}_l$ using crossover and mutation operators.
- *Update*z^*, $\forall j = 1, \dots, m$, if $z_j^* > f_j\ (\mathbb{Y})$, then set $z_j^* = f_j\ (\mathbb{Y})$.
- *Update* neighboring solutions: For each index $j \in B(i)$, if $g^{te}(\mathbb{Y}|\lambda_j, z^*) \leq g^{te}(\mathbb{P}_j|\lambda_j, z^*)$, then *set*$\mathbb{P}_j = \mathbb{Y}$ and $f_j = f(\mathbb{Y})$.
- *Update* EP: *Remove* from EP all vectors dominated by $f(\mathbb{Y})$.

Insert$f(\mathbb{Y})$ to EP if no vector in EP dominate $f(\mathbb{Y})$.

Step 3 – Termination rule

- If $t = t_{max}$, then *stop* and *output* EP,
  else $t = t + 1$, *goto* Step 2.

3 The proposed MOEA based evolutionary clustering model

Mathematically, a network is modeled as graph of pair-wise edges between its nodes. Assuming, for example, a friendship graph G modeling a social network $\mathcal{N}$, the pairwise friendship connections between individual entities of $\mathcal{N}$ can be modeled by the pair (V, E). The set of n individuals or entities in $\mathcal{N}$ is denoted as the set of nodes or vertices $V = \{v_1, v_2, \dots, v_n\}$ in G while the friendship connection between any pair of individuals in $\mathcal{N}$ is denoted as edge (v_i, v_j) in E, i.e. $E = \{(v_i, v_j) | 1 \leq i, j \leq n \wedge i \neq j\}$. Normally, any undirected graph G can be represented by an adjacency matrix A. Rows and columns of A are labeled with the vertices of V and the entry (i, j) is 1 if vertex v_i is adjacent to vertex v_j, i.e. if $(v_i, v_j) \in E$. In list notation, matrix A can be represented by a set of n adjacency lists $L = \{l_1, l_2, \dots, l_n\}$, one list l_i for each vertex $v_i \in V$ aggregating all 1 entries in row i. Thus, $|l_i| = \sum_{j=1}^{n}(i, j)$ and $|L| = \sum_{i=1}^{n}|l_i|$. Mathematically denoted, n is said to be the cardinality of G, $|l_i|$ is said to be the degree of vertex v_i, while $|L|$ is said to be the volume of G.

The proposed multi-objective evolutionary clustering algorithm, $MOEC$, is based on $MOEA/D$ [10]. Assume that $\mathbb{G} = \{G^1, G^2, \dots, G^T\}$, be a network captured at T time steps. The first decision step is to select a proper and efficient chromosome representation. The adopted representation is the locus-based adjacency representation being proposed by Park and Song [11]. In locus-based representation, each chromosome $P^t = \left(P_1^t, P_2^t, \dots, P_{n^t}^t\right) \in \mathbb{P}^t$ is represented as a fixed-length vector of n^t genes where n^t is the total number of nodes in the network at time step t (i.e. n^t is the cardinality of G^t). The genotype-phenotype decoding function δ of individual P^t will outline the community structure of the network $\mathbb{G}$ at time step t, i.e. $\delta(P^t)$: $\mathcal{C}^t = \{C\}_{i=1}^{K}$.

Given that $MOEA/D$ is population-based optimization algorithm, the next step, then, is to create an initial population $\mathbb{P}^t$ of $PopSize$ chromosomes, i.e. $\mathbb{P}^t = \{P_1^t, P_2^t, \dots, P_{PopSize}^t\}$. The traditional approach to create an initial population is to randomly distribute initial chromosomes in the search space. The next step is to formulate the objective functions. The formulation is suggested to be based on ID as the first objective function (Eq. 7), while the second objective (Eq. 8) is formulated as the inverse of NMI.

$$F1 = min\ ID(\mathcal{C}) = \sum_{i=1}^{K} 1 - \frac{m_i}{n_i(n_i-1)/2} \quad (7)$$

$$F2 = min(1 - NMI) \quad (8)$$

The whole process of $MOEA/D$ for the multi-objective evolutionary clustering problem ($MOEC$) of $\mathbb{G}$ at time step t can be described as an iterative evolution function Ψ^t: $\{\mathbb{P}^t, EP^t\} \rightarrow \{\mathbb{P}^{t\prime}, EP^{t\prime}\}$, where Ψ^t is a compositional generation-wise operator combining a sequence of genetic-like operations including parent selection, crossover, mutation or heuristic migration, and maintaining of near-Pareto archive. The population and external archive at time step t starts with an initial random population $\mathbb{P}_0^t$ and empty archive, $EP_0^t = \{\}$, and continue evolution until a maximum number of iterations $iter_{max}$ has been reached. Uniform crossover and mutation operators are used with probability p_c and p_m, respectively. Consider two individuals P_1^t and P_2^t to be the two participating parents in the crossover. A child $P_i^{t\prime}$ can be formally generated by: $\forall 1 \leq j \leq n$

$$P_{i,j}^{t\ \prime} = \begin{cases} P_{1,j}^t\ if\ r \leq 0.5 \\ P_{2,j}^t\ otherwise \end{cases} \quad (9)$$

where $r \sim [0,1]$ is a uniform random number. For the mutation operator, the value of the mutated gene $P_{i,j}^t$ can be selected to be any value v belongs to its neighbors, i.e. $A\left(P_{i,j}^t, v\right) = 1$.

Additionally, the *heuristic migration* operator proposed in [12] with probability of occurrence p_h is also adopted in this work to replace mutation operator. This operator is proposed to act as a heuristic partition generator that can exploit information from the neighborhood relations between nodes of the network. For an individual P_i^t and under the control of p_h, the heuristic migration operator will change the community belongingness of node j, i.e. $P_{i,j}^t$ if it appears to be either weakly- or neutrally-neighborhood node with all other nodes belong to the same community. If $P_{i,j}^t$ is seem to be a weakly-neighborhood node in community C, then the migration operator will migrate it to another community that would satisfy with its nodes the highest strongly-neighborhood relation. Otherwise if $P_{i,j}^t$ is a neutrally-neighborhood node in community C, then the migration operator will either leave the node

inside its community or migrate it to another community that would also satisfy inside it an equal neighborhood relation.

4 Experimental results

In this section, we will test the performance of the proposed $MOEC$ framework (denoted in the results by its snapshot function ID) against Folino and Pizzuti's model $MOEC$ framework (denoted as Q) [3]. The characteristic components of $MOEA/D$ are quantified to the following.$PopSize = 100$, $s = 5$, $iter_{max} = 100$, and $p_c = 0.8$. Also, the results showed the impact of the heuristic migration operator on the final performance of the competent $MOEC$ models. Either mutation operator with $p_m = 0.2$ or heuristic migration operator with $p_h = 0.5$ is used. The performance of the competent models is evaluated in terms of convergence reliability, where average NMI over ten different runs for each network and at each time step t (denoted as $\overline{NMI}(\mathcal{C}_t,\mathcal{C}_t^*)$).

To make a controlled check of how well the proposed model performs against other state-of-the-art $MOEC$ models, it will be wise to experiment them with a set of computer generated networks of different complexity levels. In this study, the benchmarks of Kim and Han [10] are used. Two benchmarks composed of a fixed number of communities and two other benchmarks are divided into a variable number of communities. Each benchmark has 10 time steps to evolve its original network. The first two benchmarks have 128 nodes divided into 4 communities with 32 nodes per community. Each node has connection density equals to 16 and shares λ (equals to 3 or 5) inter-connections. These networks are denoted as $\mathbb{Net}_{\mathbb{FIX}_3} = \left\{\mathcal{N}et_{\mathcal{FIX}_{3_t}} | 1 \leq t \leq 10\right\}$ and $\mathbb{Net}_{\mathbb{FIX}_5} = \left\{\mathcal{N}et_{\mathcal{FIX}_{5_t}} | 1 \leq t \leq 10\right\}$. The second two benchmarks are obtained by modifying the generation method of $\mathbb{Net}_{\mathbb{FIX}}$ to introduce the forming and dissolving of communities and the attaching and detaching of nodes. The initial network contains 256 nodes, divided into 4 communities and 64 nodes per community. From each community, 8 nodes are selected randomly and a new community is generated from the selected 32 nodes. This is repeated for 5 timestamps, then the nodes return back to their original communities. The average degree of each node in a cluster is set to the half of the size of this cluster. Furthermore, at each time step 16 nodes are randomly deleted and 16 new nodes are added to the network. The networks are denoted, respectively, by $\mathbb{Net}_{\mathbb{VAR}_3}$ and $\mathbb{Net}_{\mathbb{VAR}_5}$. Table-1 and 2 report the comparison results of these networks. The result corresponding to the winner model is given in bold at each time step. Moreover, Figure 2 – 5 qualitatively depict performance comparison. The results reported in the tables and figures clearly reflect the ability of the proposed $MOEC$ (based on ID) to beat modularity based $MOEC$ in almost all test cases, except in one case see Figure- 5 right graph. The success of the proposed $MOEC$ model over Folino and Pizzuti's $MOEC$ model can be noticed at almost all time steps, including the first time step. This says that the proposed model has the ability to satisfy a two-fold goal. The first goal is to get a more accurate detection in the structure of the communities of a static network (refer to the results of the networks at time step $t = 1$). Here, one can see that ID model outperforms Q model in the accuracy of detection. The second goal is that the proposed model has the ability to track the changes in the structure of the communities (refer to the results of the networks at time step $t > 1$) more accurate than Folino and Pizzuti's model $MOEC$. In other words, one can say that both ID and NMI has a better collaboration activity than both Q and NMI. For Figure-5 (right graph), one can say that the collaboration between Folino and Pizzuti's model and the heuristic operator got more clear and advantageous than the collaboration with ID model

Table 1- Performance comparison of $MOEC$ based on modularity model against our model on Kim and Han [10] networks. Both models are tested with no heuristic, i.e. only traditional mutation with $p_m = 0.2$ is used.

Network	*t*	1	2	3	4	5	6	7	8	9	10
Net_FIX_3	*Q*	0.0219	0.1192	0.247	0.398	0.4803	0.476	0.4763	0.48529	0.511	0.51713
	ID	**0.3611**	**0.5141**	**0.6049**	**0.668**	**0.694**	**0.744**	**0.7166**	**0.73276**	**0.744**	**0.7555**
Net_FIX_5	*Q*	0.0074	0.0773	0.1187	0.2204	0.222	0.271	0.2897	0.2638	0.276	0.3323
	ID	**0.1573**	**0.2236**	**0.3114**	0.3453	**0.312**	**0.354**	**0.3394**	**0.3575**	**0.361**	**0.3781**
Net_VAR_3	*Q*	0.6213	0.6296	0.5928	0.606	0.508	0.588	0.7019	0.7074	0.744	0.7760
	ID	**0.6549**	**0.7703**	**0.7999**	**0.7991**	**0.704**	**0.743**	**0.8186**	**0.8484**	**0.898**	**0.9330**
Net_VAR_5	*Q*	**0.5908**	0.561	0.5354	0.5117	0.393	0.398	0.5123	0.5326	0.557	0.5881
	ID	0.5831	**0.6027**	**0.5942**	**0.5586**	**0.453**	**0.474**	**0.5714**	**0.5865**	**0.603**	**0.6388**

Table 2- Performance comparison of $MOEC$ based on modularity model against our model on Kim and Han [10] networks. Both models are tested with heuristic migration operator undrer $p_h = 0.5$.

Network	*t*	1	2	3	4	5	6	7	8	9	10
Net_FIX_3	*Q*	0	0.423	0.621	0.712	0.747	0.756	0.767	0.775	0.853	0.856
	ID	**0.299**	**0.842**	**0.899**	**0.947**	**0.972**	**0.974**	**0.968**	**0.972**	**0.964**	**0.977**
Net_FIX_5	*Q*	0	0.024	0.037	0.050	0.067	0.085	0.131	0.155	0.204	0.279
	ID	**0.188**	**0.340**	**0.543**	**0.636**	**0.4300**	**0.620**	**0.621**	**0.670**	**0.646**	**0.674**
Net_VAR_	*Q*	**0.836**	**0.936**	0.932	0.920	0.859	0.853	0.931	0.951	0.971	0.979
	ID	0.655	0.930	**0.964**	**0.987**	**0.940**	**0.944**	**0.999**	**0.996**	**0.997**	**0.995**
Net_VAR_5	*Q*	**0.588**	**0.937**	**0.875**	**0.859**	**0.776**	**0.774**	**0.854**	**0.888**	**0.903**	**0.947**
	ID	0.562	0.622	0.666	0.687	0.609	0.672	0.805	0.853	0.870	0.917

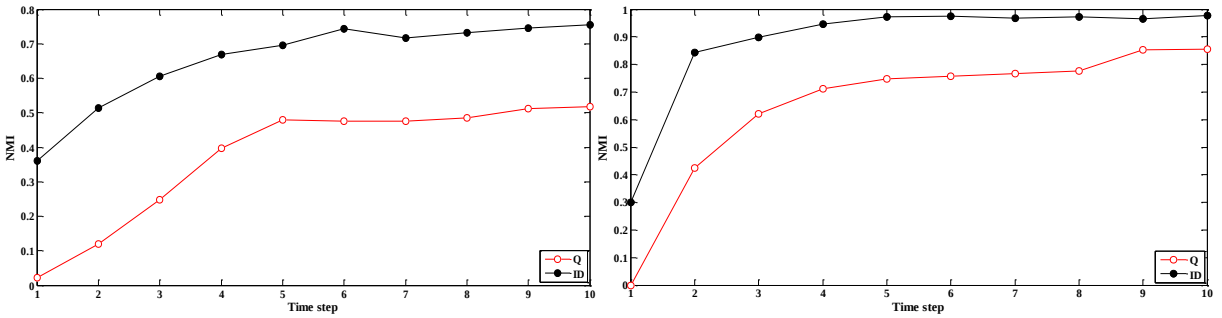

Figure 2- Performance comparison of MOEC based on modularity (line with filled circle) model against our model (line with unfilled circle) on $\mathbb{Net\_FIX\_3}$ (left: with no heuristic, i.e. $p_m = 0.2$), (right: with heuristic, i.e. $p_h = 0.5$).

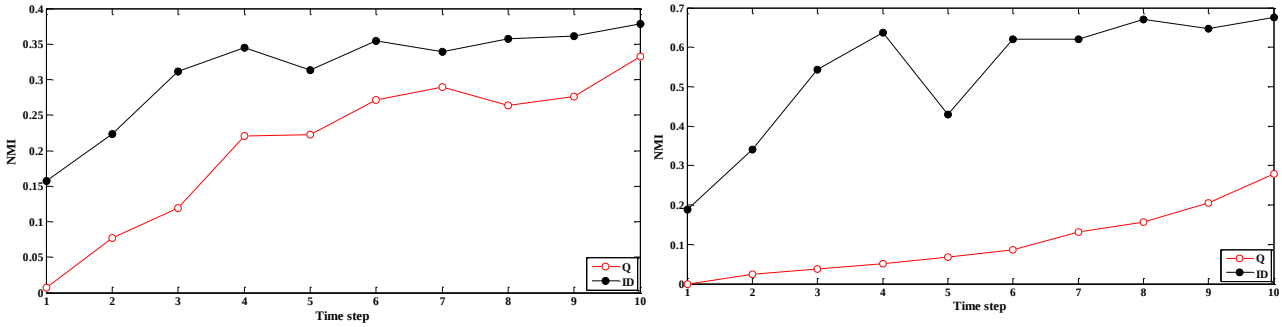

Figure 3- Performance comparison of MOEC based on modularity (line with filled circle) model against our model (line with unfilled circle) on $\mathbb{Net\_FIX\_5}$ (left: with no heuristic, i.e. $p_m = 0.2$), (right: with heuristic, i.e. $p_h = 0.5$).

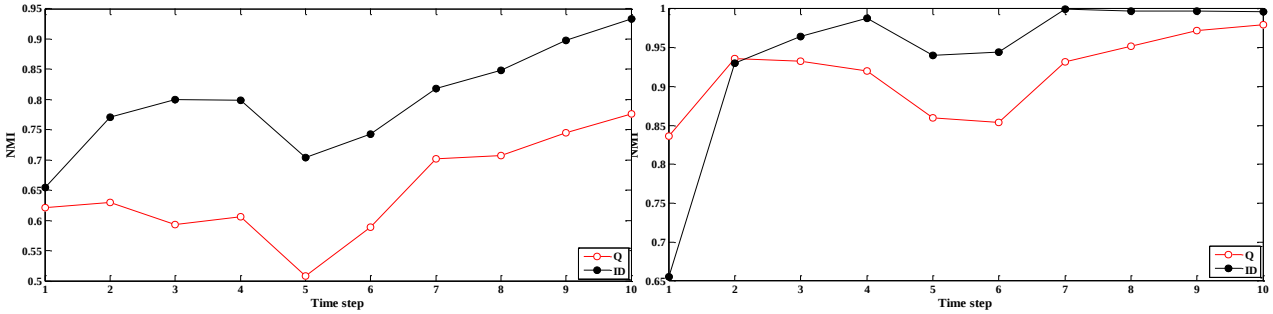

Figure 4- Performance comparison of MOEC based on modularity (line with filled circle) model against our model (line with unfilled circle) on $\mathbb{Net\_VAR\_3}$ (left: with no heuristic, i.e. $p_m = 0.2$), (right: with heuristic, i.e. $p_h = 0.5$).

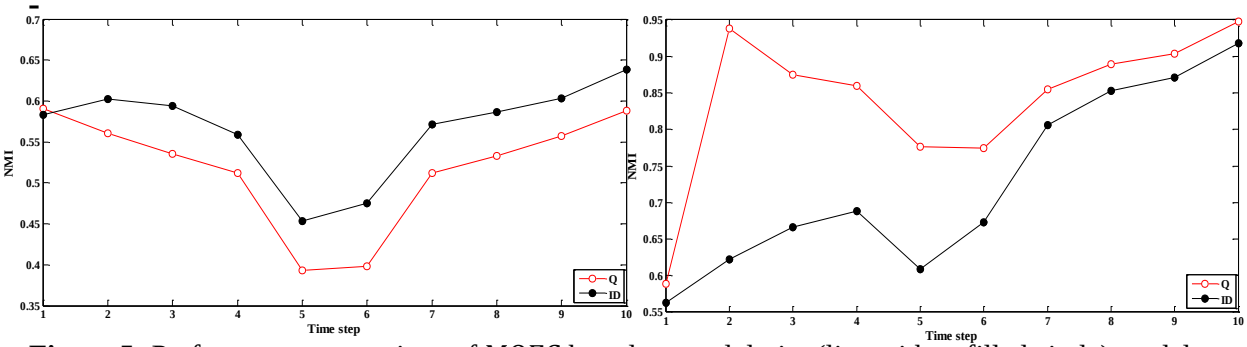

Figure 5- Performance comparison of MOEC based on modularity (line with unfilled circle) model against our model (line with filled circle) on $\mathbb{Net\_VAR\_5}$ (left: with no heuristic, i.e. $p_m = 0.2$), (right: with heuristic, i.e. $p_h = 0.5$).

5 Conclusions

In fact, the detection of the community structure in the complex networks that have dynamic behavior encounters many challenges. Community detection in dynamic networks is normally known in literature as evolutionary clustering. In evolutionary clustering, both snapshot cost and temporal cost should be considered to tackle the problem. In this paper, a new snapshot cost, based on internal density of the community, is formulated and together with the temporal cost (being signified by NMI) are utilized to form a multi-objective evolutionary clustering framework. By comparing the performance of the proposed model with the state-of-the-art modularity based model, the results reveal more accurate results. Another ramification to the current work could be stated by investigating other temporal cost alternatives or formulations.

References

1. Chakrabarti, D. Kumar, R. & Tomkins, A. **2006**. Evolutionary clustering. in: Proceedings of the 12th ACM International Conference on Knowledge Discovery and Data Mining (KDD'06), pp. 554–560.

2. Kim, M.S.، & Han, J. **2009**. A particle-and-density based evolutionary clustering method for dynamic networks. *Proceedings of the VLDB Endowment*, 2(1), 622-633.

3. Folino, F. & Pizzuti, C. **2014**, An evolutionary multiobjective approach for community discovery in dynamic networks. *Knowledge and Data Engineering. IEEE Transactions on*, *26*(8), 1838-1852.

4. Folino, F.، & Pizzuti, C. **2010**. A multiobjective and evolutionary clustering method for dynamic networks. In Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2010 International Conference on (pp. 256-263).

5. Newman, M. E. J. and Girvan, M. **2004**. Finding and evaluating community structure in networks. *Physical Review E*, vol. 69, no. 026113.

6. MacKay, D. J. C. **2002**. *Information theory, Inference and Learning Algorithms*. Cambridge, U.K.: Cambridge University Press.

7. Radicchi, F. Castellano, C. Cecconi, F. Loreto, V. & Parisi, D.**2004**. Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. USA*, vol. 101, no. 9, pp. 2658–2663, 2004.

8. Fortunato, S.، **2010**. Community detection in graphs. *Physics Reports*, *486*(3–5), 75–174.

9. Coello، C.A.C. Van، D.A., & Lamont, G.B. **2002**. *Evolutionary algorithms for solving multi-objective problems* (Vol. 242). New York: Kluwer Academic.

10. Zhang, Q. & Li, H. **2007**. MOEA/D: A multi-objective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731.

11. Park، Y. J. & Song، M. S. **1998**. A genetic algorithm for clustering problems, *Proc. 3rd Annu. Conf. Genet. Algorithms*, pp. 568–575.

12. Attea, B.A. W.A. & Hariz, M.F. Abdulhalim. **2016**. Improving the performance of evolutionary multi-objective co-clustering models for community detection in complex social networks. *Swarm and Evolutionary Computation*, vol. 26, pp. 137-156.