

## MODM: multi-objective diffusion model for dynamic social networks using evolutionary algorithm

Iram Fatima · Muhammad Fahim ·  
Young-Koo Lee · Sungyoung Lee

Published online: 7 May 2013  
© Springer Science+Business Media New York 2013

**Abstract** A lot of research efforts have been made to model the diffusion process in social networks that varies from adoption of products in marketing strategies to disease and virus spread. Previously, a diffusion process is usually considered as a single-objective optimization problem, in which different heuristics or approximate algorithms are applied to optimize an objective of spreading single piece of information that captures the notion of diffusion. However, in real social networks individuals simultaneously receive several pieces of information during their communication. Single-objective solutions are inadequate for collective spread of several information pieces. Therefore, in this paper, we propose a Multi-Objective Diffusion Model (MODM) that allows the modeling of complex and nonlinear phenomena of multiple types of information exchange, and calculate the information worth of each individual from different aspects of information spread such as score, influence and diversity. We design evolutionary algorithm to achieve the multi-objectives in single diffusion process. Through extensive experiments on a real world data set, we have observed that MODM leads to a richer and more realistic class of diffusion model compared to a single objective. This signifies the correlation between the importance of each individual and his information processing capability. Our results indicate that some individuals in the network are naturally and significantly better connected in terms of receiving information irrespective of the starting position of the diffusion process.

---

I. Fatima · M. Fahim · Y.-K. Lee (✉) · S. Lee  
Department of Computer Engineering, Kyung Hee University, Seocheon-dong, Giheung-gu,  
Yongin-si, Gyeonggi-do 446-701, Korea  
e-mail: [yklee@oslab.khu.ac.kr](mailto:yklee@oslab.khu.ac.kr)

I. Fatima  
e-mail: [iram.fatima@oslab.khu.ac.kr](mailto:iram.fatima@oslab.khu.ac.kr)

M. Fahim  
e-mail: [fahim@oslab.khu.ac.kr](mailto:fahim@oslab.khu.ac.kr)

S. Lee  
e-mail: [sylee@oslab.khu.ac.kr](mailto:sylee@oslab.khu.ac.kr)

**Keywords** Multi-objective optimization · Evolutionary algorithm · Information diffusion · Dynamic social networks

## 1 Introduction

A social network is a graph representing the interactions or interconnections among individuals based on their common interests, activities, and demographic identities. Diffusion process in social networks has attracted much attention and a lot of research efforts have been made in this field from all areas of academic interest, such as physics, mathematics and computer science [1–3]. It is widely believed that user-to-user exchanges, also known as “word of mouth,” can spread contents, ideas, or information widely and quickly throughout the network [4, 5]. The fundamental purpose of each diffusion process is to differentiate a set of the individuals on the basis of their social ability for information manipulation and propagation [6]. In most of the diffusion models the notion of ‘information’ is restricted to a single unit during the diffusion process to evaluate the importance of individuals in a social network. Conversely, in real social networks, individuals communicate their ideas and feelings in a spur of a moment with various people like family, friends, and relatives in homes, offices, and universities. So in a true social network, a variety of information like news, rumors, gossips, stories, and announcements is manipulated and spread at the same time [7].

The collective diffusion process for these various information types seems chaotic. Each piece of information<sup>1</sup> has a separate spread process according to its type, associated constraints, and importance. For example, some information is independent of any competition (e.g., TV news) while some ideas, opinions, and products compete with all the other content (e.g., product adoption, political elections) for the scarce attention of the users. Previously, researchers simulated these types of information independently in separate diffusion models as a single-objective problem to analyze a social network [8, 9]. Most of the approaches spread ‘information’ as a single unit with ‘active’ or ‘inactive’ status in order to group the individual into two categories. Therefore, at the end of the diffusion process the individuals with the ‘active’ status achieve the single objective of the diffusion, whereas ‘inactive’ individuals have no effect on the diffusion process [7, 8]. However, it is intractable to distinguish among ‘active’ individuals in order to find the differences between them according to their network property. Furthermore, in a real social network people do not lie between two statuses of either ‘active’ or ‘inactive’ to show their significance in the network. Instead, more granularity of individual importance is required to find the differences between them that can reflect their information propagation capability in the network. The situation becomes more complex for single-objective diffusion models when individuals propagate multiple types of information. At present, none of the existing diffusion models are able to comprehensively handle the aforementioned problems. Therefore, in this study, we first formulate the diffusion process as a multi-objective

---

<sup>1</sup>We use information and piece of information interchangeably throughout the paper.

optimization problem to model the information spread closer to a real social network. We propose a multi-objective diffusion model (MODM) that assigns a value of importance to each individual according to his information manipulation and propagation ability. The goal of MODM is to selfishly maximize the amount of information possessed by each individual during communication. The key difference from earlier studies is that we achieve multiple objectives in terms of diverse information spread and calculation method to measure the propagation capabilities of individuals.

Our contribution in this work is fourfold. First, we propose a novel multi-objective diffusion model based on evolutionary algorithm that is superior in terms of achieving more than one objective at a time in a social network. Second, we introduce a method to propagate the multiple types of information (i.e., independent, mutually exclusive, and competing) in binary schema formats. This gives the flexibility in terms of modeling the set of information with variations in bits of string according to their types and constraints. Third, evolution fitness criterion is designed in order to calculate the multi-objective score (value of importance) of each individual. We believe that different aspects of information transmission are important to determine which individuals have high worth in multiple types of information spread. Therefore, the proposed evolution fitness criteria are based on information score, influence, and diversity. Fourth, we maintain an information history log for each individual to keep track of information flow during the whole diffusion process. This facilitates the evolution fitness criteria to calculate the true information value of each individual during their interactions. Our proposed MODM, to achieve multiple objectives by propagating a set of information with various evaluation criteria, to the best of our knowledge has never been applied before. We perform extensive experiments to validate the advantages of MODM on a real-world data set. The solution given by our model, more realistically reveals the modeling of complex and nonlinear phenomena of information exchange to affect the total information worth of each individual. The experimental results show superiority of MODM over single information propagation and single evaluation criterion.

The rest of the paper is organized as follows. We briefly describe related works and their limitations in Sect. 2. In Sect. 3, we formulate the multi-objective diffusion problem. Section 4 presents our proposed multi-objective diffusion model. In Sect. 5, we analyze our experimental results to validate the effectiveness of MODM. Finally, the conclusion and future work are presented in Sect. 6.

## 2 Related work

Smart models of information diffusion and influence maximization have been studied in many different forms, for example, the transmission of political opinions and news in political science [10], the diffusion of innovations in management science [11], the value of novel information in organizational behavior [12], and the propagation of obesity and smoking behaviors in public healthcare [13]. In the past, several models have been introduced to simulate information diffusion through a network [14–17]. The linear threshold model [14] and independent cascade model [15, 16] are the most widely used diffusion methods. In the linear threshold model [14] a node is influenced

by each neighbor according to a given weight. Each node chooses a threshold value uniformly at random from interval  $[0, 1]$ ; this represents the weighted fraction of node neighbors that must become active in order for a node to become active. Given a random choice of threshold and an initial set of active nodes, the diffusion process unfolds deterministically in discrete steps. In step ' $t$ ', all nodes that are active in step ' $t - 1$ ' remain active and activation of any node depends on the total weight of its active neighbors that must be above threshold value. In the independent cascade model [15, 16], the diffusion process starts with an initial set of active nodes and the process unfolds in discrete steps according to the following randomized rule. When a node first becomes active in step ' $t$ ', it is given a single chance to activate each currently inactive neighbor. It succeeds according to a probability value. If a node succeeds then its neighbors become active in step ' $t + 1$ '. Regardless of whether or not a node succeeds, it cannot make any further attempts to activate its neighbors in subsequent rounds. However, it requires the parameters that represent propagation probabilities through links to be specified in advance.

A genetic algorithm based diffusion model [18] was introduced to combine the advantage of a genetic algorithm paired with a form of Holland synthetic hyperplane-defined objective functions for a parameter-free diffusion process. Chromosomes in the genetic algorithm represented individuals in a network, and the crossover operator modeled the interactions between them. Information diffusion proceeded in the crossover operation in all timestamps. During a crossover process, the tail of a chromosome containing some information is replaced when the sum of information is lower than that of the new tail. This causes the original information on the chromosome to be totally lost, although it may not conflict with the new incoming information. The information value of each individual is calculated by adding the score of each gene within a chromosome.

In order to create realistic models for diffusion process, it is important to train with a true picture of the social interactions between individuals and the parameters that affect the propagation process. In the cases of independent cascade and linear threshold models, two kinds of data, a social network and probabilities to the edges capturing the degree of influence among individuals, are required. For example, if edge  $(v, u)$  has 0.45 probability that user ' $v$ ' influences ' $u$ ', then  $v$ 's action will propagate to ' $u$ ' with a fixed probability. In real life, edge probabilities are not available for a social network, so previous work either makes assumptions about these probabilities or uses other heuristics to calculate them. This poses a big problem of estimating this probability from the real data. GADM [18] proposed a parameter-free diffusion process with the help of genetic algorithm and Holland hyperplane-defined function. However, the methods described above use the diffusion process to achieve single objective according to their domain of interest. That is why these methods are not appropriate to find the optimized diffusion solution for more than one objective under single diffusion process. Therefore, we propose a multi-objective diffusion model (MODM) to achieve a multi-objective optimized diffusion process that represents multiple types of information propagation with evolution fitness criteria to make the diffusion process closer to real social networks.

### 3 Multi-objective diffusion in social networks

A social network is illustrated as a graphical representation of interactions between a set of vertices. Some famous social networks include online social networks, where vertices are user accounts and edges represent friendships among accounts. Similarly, in communication networks, vertices represent e-mail addresses or telephone numbers and edges represent e-mails sent or telephone calls with the time of interaction. A typical social network tends to expand over time, with newly added nodes and edges being incorporated into the existing graph with time intervals.

**Definition 1** A dynamic social network  $G = (g_1, g_2, \dots, g_T)$ , is a directed multi-graph, where  $g_i = (V_i, E_i)$  represents the bag of vertices  $V_i$  and edges  $E_i$ , at a particular time interval  $t_i \in T$ . A node  $v \in V_i$  shows an individual and an edge  $(u, v) \in E_i$  represents an interaction between two individuals during their communication.

The information propagation in any social network depends on the type of a diffusion model. A diffusion model accepts as input a graph structure and state of every individual at a time ' $t$ '. It returns a new state of the individual on time ' $t + 1$ ' according to its interaction with other individuals. The process continues until all the interactions between individuals are exhausted. The conventional diffusion models can be roughly divided into two categories: (1) an independent cascade model [15], and (2) a linear threshold model [14]. In both models, the diffusion process can be regarded as a single-objective optimization problem  $(\Omega, D)$  as described below.

**Definition 2** Single-objective diffusion model determines a set of individuals  $I^*$  for which

$$D(I^*) = \max_{I \in \Omega} D(I) \quad (1)$$

where  $\Omega$  is the the a unit of information propagated among all individuals  $I$  in the network  $G$ , and we assume  $D$  to be the objective function for optimization.

The single-objective diffusion models have been widely applied, as most conventional diffusion processes [14–17] are based on this single-objective optimization problem. However, they have several disadvantages. (1) The single-objective diffusion models attempt to solve the problem of diffusion in unitary format to fulfill a single criterion and thus optimize a network on one direction. (2) The diffusion process based on a single objective may fail to maintain the monotonicity property of information during individuals' interaction. (3) Many single-objective algorithms require some prior information about the influence of vertices in the form of edge weights in the network; this influential information is mostly missing in real world networks. (4) A single-objective optimization cannot optimize the multiple types of information on one evaluation criterion. (5) A single-objective diffusion model returned by single-objective algorithms may not be suitable for networks with multiple potential diffusion measures. The difficulty in selecting an appropriate criterion in single-objective diffusion model can be handled using a more natural approach that considers the diffusion process as a multi-objective optimization problem which can be defined as follows.

**Definition 3** Multi-objective diffusion model determines a set of individuals  $I^*$  for which

$$D(I^*) = \max_{I \in \Omega} (D_1(I), D_2(I), \dots, D_m(I)) \quad (2)$$

where  $\Omega$  is a set of multiple types of information (e.g., news, gossips, rumors, and reports) and  $m$  is the number of objective functions for evolution fitness criteria. In the above equation  $D_i$  represents the  $i$ th objective function of multi-information. Compared to the single-objective diffusion process, the multi-objective diffusion process has the following advantages.

- The optimal solution obtained by the multi-objective diffusion process defined by  $(\Omega, D_1, \dots, D_m)$  always contains the optimal solutions of the single-objective diffusion process defined by  $(\Omega, D_1), \dots, (\Omega, D_m)$ .
- The multi-objective diffusion process can always find a set of individuals that are as good as or better than those of the single-objective diffusion process. Most specifically, in some situations where the best solution corresponds to a trade-off between different objectives, only the multi-objective diffusion process is able to find it.
- The multiple objectives can measure characteristics of a diffusion process from various perspectives, and thus avoid the risk that a single objective may only be suitable to a certain kind of information spread. Moreover, the multi-objective optimization process achieves a balance of the multiple objectives and can effectively avoid being trapped in a single dimension of the diffusion process.
- The multi-objective diffusion model reveals network analysis from different angles, which help to discover complex and comprehensive information diffusion trends in social network.

#### 4 The proposed multi-objective diffusion model

For the multi-objective diffusion problem, both mathematical programming and heuristic approaches can be applied to solve it. In this paper we designed the Genetic Algorithm (GA), a type of heuristic approach, to solve the problem of multi-objective diffusion. Compared to mathematical programming techniques, GA has many advantages [19, 20], such as simultaneously generating a set of candidate solutions and easily dealing with a discontinuous and concave solution [19]. Conventional evolutionary multi-objective algorithm is designed for numerical optimization problems. When it is applied to the multi-objective diffusion, many components of GA need to be redesigned. This is not a trivial task, because the design of these components directly determines the desired output in terms of algorithm performance.

Concretely, the multi-objective diffusion process with GA faces the following challenges: (1) Modeling of multiple types of information, it is a set of information with different adoption and diffusion criteria, and (2) selection of optimization evaluation criteria. The fitness functions should reflect the semantic characteristics of individuals from different aspects, such as score, influence and diversity. Ideal fitness functions have to better contain intrinsic conflicts, such that the optimal set of individuals could be obtained through the trade-off of multiple evaluation criteria. (3) For

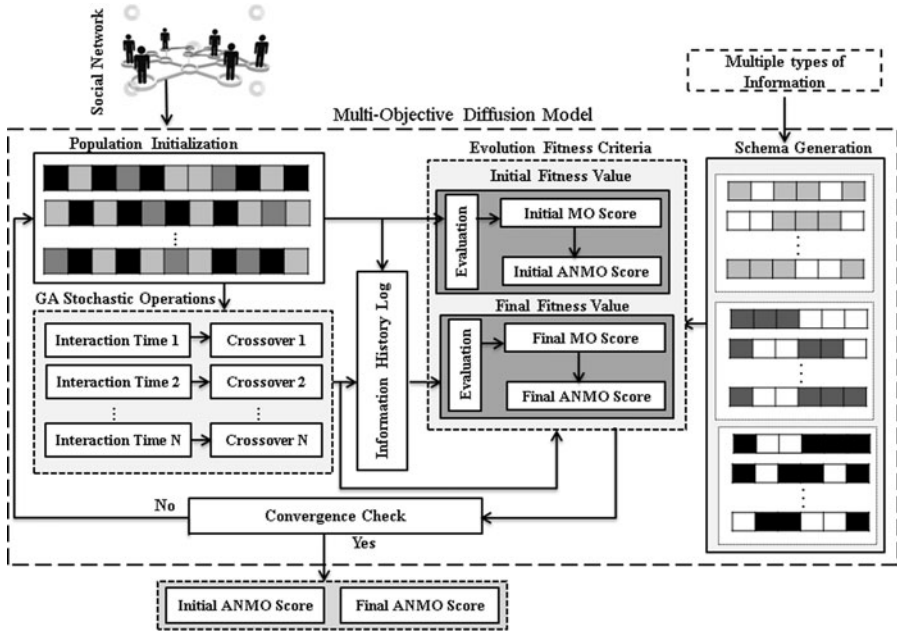


Fig. 1 The proposed architecture of MODM

effective genetic representation, the genetic representation should be delicately designed according to the characteristics of the diffusion process, since it determines the algorithm performance and scalability to a large extent. We now show how the multi-objective GA with binary string chromosomes and one-point crossover can be used as the framework for a multi-objective diffusion model. For this purpose, our proposed architecture is represented in Fig. 1. It has six major modules: (a) multiple types of information, (b) schema generation, (c) population initialization, (d) information history log, (e) evolution fitness criteria, and (f) GA stochastic operations. The details of each component are described in subsequent sections, and the pseudocode of MODM is shown in Algorithm 1.

#### 4.1 Multiple types of information

Social networks empower and inspire multiple types of information with separate diffusion processes based on their characteristics. In MODM we propagate three kinds of information: (a) independent information, (b) mutually exclusive information, and (c) competing information. The difference between these information types is as follows.

##### 4.1.1 Independent information

Independent information spreads autonomously, without any constraint, and an individual can hold many independent pieces of information. For example, information

**Algorithm 1:** Multi-objective diffusion model

---

**Input:**  $sng$  – social network graph  
 $mt$  – maximum trials  
 $ct$  – convergence threshold

**Output:**  $I_{ANMOS}$  – Initial average normalized multi-objective score  
 $F_{ANMOS}$  – Final average normalized multi-objective score

**Begin**

- 1  $g = load(sng)$
- 2  $nv = getUniqNodes(g)$
- 3  $edg = getUniqEdges(g)$
- 4  $ts = getUniqTStamps(g)$
- 5  $itr = 1$
- 6 **While**( $ct || isequal(mt, itr)$ )
- 7      $schNum = rand()$
- 8      $sch = SchemaGen(schNum)$
- 9     **for**  $j = 1: size(nv)$
- 10          $ininode = rand()$
- 11          $I_{pop} = concatenate(ininode, multiObjScore(ininode, sch))$
- 12     **end**
- 13      $I_{ANMOS} = fanmsCal(I_{pop}, I_{ANMOS}, itr)$
- 14     **for**  $l = 1: size(ts)$
- 15          $F_{pop} = fGAOperation(I_{pop}, sch, edg)$
- 16     **end**
- 17      $F_{ANMOS} = fanmsCal(F_{pop}, F_{ANMOS}, itr)$
- 18      $itr = itr + 1$
- 19 **end**

**End**

---

about different news is independent and can diffuse independently in the network without any spreading constraints. In the healthcare domain, information about diseases like diabetes, heart attack, stroke, and blood pressure is independent from each other. In the cellular market, information about cell phones such as Android phones, Apple phones, and Windows smart phones is also independent.

#### 4.1.2 Mutually exclusive information

Unlike independent information, an individual can hold only one piece of information from a set of mutually exclusive information. Upon selecting a piece of information from mutually exclusive set, he automatically denies other pieces of information from the rest of the set. For example, two music concerts are going to be held at the same time (i.e., mutually exclusive), and an individual can choose only one to attend. Upon selecting one concert, he denies the other. In the healthcare domain a patient can choose between surgery or laser therapy (mutually exclusive) to cure his illness. In the cellular market a user can choose between postpaid or prepaid connection with his contact number.



### 4.1.3 Competing information

Similarly to mutually exclusive information, an individual can hold only one piece of information from a list of competing information. However, competing information can be updated with certain constraints. For example, two music concerts are going to be held at the same time, and an individual can attend only one. One concert is popular among people, so when an individual knows both pieces of information he would update his choice to the most popular one and inform others about it. In the healthcare domain, usage of antibiotics is updated according to their ranking in different seasons, side effects and environments. In the cellular market a user can update his data plan according to recent cost effective and usability offers.

## 4.2 Schema generation

Schema is generated using well-known Holland's hyperplane defined functions (HDFs) [21]. Each schema is a set of binary values that gives flexibility in terms of modeling multiple types of information with variations in bits of the schema, where '\*' represents "don't care" terms that are either zero or one. Schema generation begins from simple binary strings and becomes more complex to second and third levels by combing previous levels. Each schema string has a start position, length, encoding and score which make one schema unique from others. The sample schema of each information type is shown in Table 1. The encoding of independent information has no constraints, while mutually exclusive information is a set of identically scored information with the same start and length positions. However, the encoding for each piece of mutually exclusive information is different, so an individual can hold only one piece of information from the whole set. Once an individual receives any mutually exclusive information, he cannot update it. In the case of competing information, start and end positions are same with different objective score and encodings. From a set of competing information an individual can choose only one piece of information at a time that can be replaced according to its score in later timestamps.

## 4.3 Population initialization

In the GA population, each individual is represented with a binary chromosome of length  $\beta$  to characterize its state during the diffusion process. This can be initially

**Table 1** HDF based schema generation

Information type	Start position	Length	Encoding	Score
Independent	4	12	**11*0001*11	0.23
	25	9	11**001*0	0.15
Mutually exclusive	17	7	1*00*01	0.14
	17	7	1*10*01	0.14
Competing	15	10	0111**1100	0.26
	15	10	0110**0100	0.36

Locus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Value	0	0	1	1	1	1	1	1	0	0	0	1	0	0	1	0	1	1	0	1

$\leftarrow I_1 \rightarrow$ 
 $\leftarrow I_2 \rightarrow$

**Fig. 2** A representative chromosome of length  $\beta = 20$

chosen according to some random distribution. Each chromosome is a set of  $n$  pieces of information, where each piece of information is represented by a short binary string that is used to indicate its spread in a network. Depending on the initial state string, each individual in the network knows certain types of information. A vector  $(x, y, l)$  is used to describe a piece of information, where  $x$  is its start-point on a chromosome,  $y$  is its score in the form of a real value between  $[0, 1]$ , and  $l$  is the length of the information. If the length of the chromosome is  $\beta$ , then  $x + l \leq \beta, x \in [1, \beta]$ . The sample chromosome of length  $\beta = 20$  is shown in Fig. 2 and contains two pieces of information,  $I_1(2, 0.25, 7)$ , and  $I_2(12, 0.45, 8)$ . If a chromosome contains an encoding of information, we say that the corresponding individual carries the corresponding information. An objective value of a chromosome is the sum of all the scores of information it contains. The process of initialization is shown in Algorithm 1, lines 9 to 12.

#### 4.4 Information history log

In addition to the score assigned by a schema to each information type, other important aspects in the calculation of an individual’s information worth are the diversity and influence of information. Information diversity is measured by the types of information retained by an individual in total, whereas the influence is determined by the frequency of the information generated for each individual during the diffusion process. For example, if someone receives the same information multiple times, the chances of adaptation for that information are high compared to the information received fewer times from the same neighbors. To keep track of the information flow from time ‘ $t_1$ ’ to ‘ $t_n$ ’, information history log is maintained that records information identity, type, score, and influence. After each information exchange among individuals, the information history log is updated according to the new incoming information and influence frequency of the existing ones as shown in Algorithm 2. This facilitates the evolution fitness criteria for calculation of an individual information worth based on the spread of numerous pieces of information.

#### 4.5 Evolution fitness criteria

The fitness criteria guide the search process to quantify the optimality of the diffusion process. Keeping the maximum amount of information as a foundational quality by assigning a relative importance to individual criteria of a fitness function is defined as follows:

$$F(x) = \sum_{i=1}^n w_i f_i(x) \tag{3}$$

**Algorithm 2:** Information history log

---

**Input:** *os* – offspring  
*sch* – schema

**Output:** *ihl* – information history log

**Begin**

```

1  for i=1: size(sch)
2  | if (os ⊆ sch[i])
3  | | if (os ⊆ ihl)
4  | | | ihl.inf = ihl.inf + sch[i].inf
5  | | | else
6  | | | | ihl.type = sch[i]
7  | | | | ihl.score = sch[i].score
8  | | | end
9  | | end
10 end

```

**End**

---

where  $x$  is an individual,  $F(x)$  is a combined fitness function,  $f_i(x)$  is the  $i$ th evaluation criterion,  $w_i$  is a constant weight for  $f_i(x)$ , and  $n$  is the total number of evaluation criteria. In order to combine multiple evaluation criteria into a scalar fitness function, a weighted sum approach is defined. Our objective is to maximize all the individual evaluation functions. In the proposed model, the evaluation criteria are score, influence and diversity of information. Score is the HDF generated value for each piece of information. Influence is maintained in the information history log by keeping a record of the number of times a piece of information was received by a particular individual. Diversity measures the total types of information retained by an individual. More specifically, the evolution fitness criterion is defined as

$$F(x) = \operatorname{argmax} [w_{scr} fScore(x) + w_{inf} fInfluence(x) + w_{dve} fDiversity(x)] \quad (4)$$

where

$$fScore(x) = \sum_{i=1}^n (score.info_i)$$

$$fInfluence(x) = \sum_{i=1}^n (influence.info_i)$$

$$fDiversity(x) = \operatorname{Count}(info.)$$

$$w_{scr} + w_{inf} + w_{dve} = 1$$

If we use constant weights in (4), the search direction in genetic algorithms becomes fixed. Therefore we propose a selection procedure with random weights to search for optimal solutions by utilizing various search directions for each evaluation criterion. In Eq. (4),  $F(x)$  is the weighted sum of all evaluation criteria that assigned multi-objective score to one individual at time  $t$ .

#### 4.6 GA stochastic operations

At each time  $t$  that a pair of individuals interacts, they exchange information, which is modeled by a crossover operation. A uniform crossover point  $c$  is selected at random from the integer range  $[1, \beta]$ . Two new state strings  $Ofsp_1$  and  $Ofsp_2$  are created by swapping the tails of interacting individuals, where the tail is defined as all positions including and following index  $c$ . For example, if the value of  $c = 2$  is selected randomly as crossover point, we exchange the bits around that point (00|101101 and 10|001011 produce 00001011 and 10101101). The idea here is that we are swapping some alleles and the resultant bit-strings are the offspring to represent the information exchange as shown in Algorithm 3.

If the edges in the social network are bidirectional, exchange of individual interaction roles is performed, and the crossover repeats. If there are multiple interactions at the same time for a single individual, one of the middle chromosome generated in each interaction is randomly chosen as a parent for crossover operation. This parent will adopt all other information existing on the other offspring and bring benefit to it. The multi-objective score of each new offspring is evaluated according to Eq. (4). If any of the offspring has a higher multi-objective score than their parents, the corresponding parent's state string is replaced in the next iteration. In the case of ties in

---

#### Algorithm 3: *fGAOperation*—GA stochastic operations

---

**Input:**  $I_{pop}$  - population  
 $sch$  - schema  
 $edg$  - interacting edges

**Output:**  $F_{pop}$  - final populations

**Begin**

```

1  crospnt = rand()
2  for i = 1: size(edg)
3    ofsp1 = concatenate (Ipop.edg[i] (crospnt: end), Ipop.edg[i+1] (1: crospnt))
4    ofsp2 = concatenate (Ipop.edg[i+1] (crospnt: end), Ipop.edg[i] (1: crospnt))
5    if (ofsp1.multiObj(sch) > ofsp2.multiObj(sch))
6      | newOfsp = ofsp1
7    else
8      | newOfsp = ofsp2
9    end
10   if (Ipop.(edg[i+1]) < newOfsp)
11     | ofsp = newOfsp
12   else
13     | ofsp = Ipop.edg[i+1]
14   end
15   Fpop = ofsp
16 end

```

**End**

---

---

**Algorithm 4:** *fanmsCal*—average normalized multi-objective score calculation
 

---

**Input:** *pop* – population  
 $P_{ANMOS}$  – Previous average normalized multi-objective score  
*itr* – iteration

**Output:**  $U_{ANMOS}$  – Updated average normalized multi-objective score

**Begin**

- 1 **for**  $j = 1$ : *size(pop)*
- 2      $Max_{score} = \max(pop)$
- 3      $Prev_{val} = P_{ANMOS}[j] / itr$
- 4      $cur_{val} = P_{ANMOS}[j] / Max_{score}$
- 5      $U_{ANMOS}[j] = Prev_{val} + cur_{val}$

6 **end**

**End**

---

the multi-objective scores of the original and an offspring, the original state string is retained as shown in lines 10 to 14 of Algorithm 3.

After the crossover operation, the multi-objective score of each individual is calculated for a particular generation. The high score of an individual could be the result of its network characteristic or based on its randomly assigned initial values. To avoid the latter bias, GA is repeated multiple times and the Average Normalized Multi-Objective score (*ANMO*) is calculated for each individual to show his information worth irrespective of the start of a diffusion process. We iteratively calculate the value of *ANMO* till the end of diffusion process. In each interaction the value of *ANMO* for each individual is updated according to his previous *ANMO* and recent multi-objective score in current population. A previous *ANMO* value is normalized with numbers of GA iterations and a new objective score is normalized with the maximum score of the network in current population. The complete workflow to calculate the *ANMO* score is shown in Algorithm 4.

The stopping criterion for GA is either a fixed number of generations or convergence to a predetermined threshold value. After performing all interactions among all the individuals the convergence test guides the MODM to continue or to stop.

## 5 Results and evaluation

This section will validate the effectiveness of MODM through experiments on a real social network. The goal of the experiments is to estimate the information value of each individual over multiple random state initializations using HDF schema and information history log. This identifies whether all individuals receive the same *ANMO* score as a result of their interaction or their score varies according to their relative position and information processing capabilities in the network.

### 5.1 Data set description

MODM is applied on publically available Enron email data set [22]. It is the large dynamic repository of e-mails of the former Enron Corporation where vertices repre-

sent e-mail addresses and directed time-stamped edges represent interaction between individuals. It has 84,716 e-mail addresses, 1,326,771 total timestamps, and 215,841 unique timestamps covering a period of approximately four years. Moreover, the out-degree and in-degree of the network are in a power-law distribution.

## 5.2 Experimental setup

We start by generating the schema for each information type with their scores. In the beginning, each individual is initialized with random binary values in chromosome. During the interaction of individuals at each timestamp initial multi-objective score is calculated. At the end of single iteration of GA, MODM processes all time-stamped edges and the final multi-objective score of each individual is normalized relative to the maximum multi-objective score in the population. We run multiple trials of GA in order to avoid any biases caused by the random assignment of initial values to individuals. At the end of diffusion process, the Average Normalized Multi-Objective score (*ANMO*) of each individual is calculated to show the relative information worth of whole population.

## 5.3 Results and discussion

In this section, we performed five experiments to show the usefulness of MODM in comparison with: (a) single-information propagation, (b) single-evaluation criteria, (c) weighted and neutral weight factor of evaluation criteria, (d) conventional network measures, and (e) an existing approach GADM [18]. In all the experiments  $x$ -axis represents the Average Normalized Multi-Objective Score (*ANMOS*) and  $y$ -axis represents the  $F_n(\text{ANMOS})$  that is defined as the proportion of individuals having *ANMO* score. The details of the experiments are as follows.

### 5.3.1 Comparison of MODM and single-information propagation

To show the significance of MODM in comparison to single-information diffusion process, we separately propagate each information type from a set of multiple types of information. The evaluation criterion during this experiment is based on Eq. (4), with equivalent weights. In Fig. 3a, the initial *ANMO* score of each information type and MODM is illustrated. The similar diffusion curve for initial *ANMO* show the similar start of diffusion process in all cases and depend on the inherent position of the network. However, for the final *ANMO* score, the diffusion curves in Fig. 3b show that MODM is better obtaining the maximum diffusion objective as compared to individual information types. The diffusion curve of mutually exclusive information is strongly clustered and shows no dispersion in the *ANMO* score maximization. The diffusion process for competing information finished quickly while independent information is better as compared to the other two information types. However, neither of them can reach maximum *ANMO* score. The result shows that MODM can be applied to model the information exchange based on a single-information type. However, the multi-objective design of the proposed model combined the benefits of multiple types of information propagation in single diffusion process that demonstrate the better information maximization during the diffusion process.

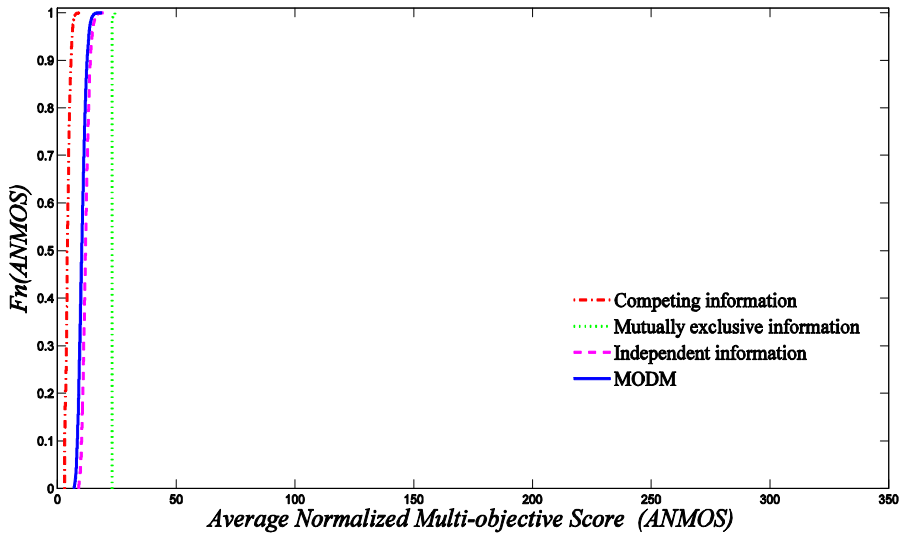


Fig. 3a Initial ANMO score for MODM and single-information types

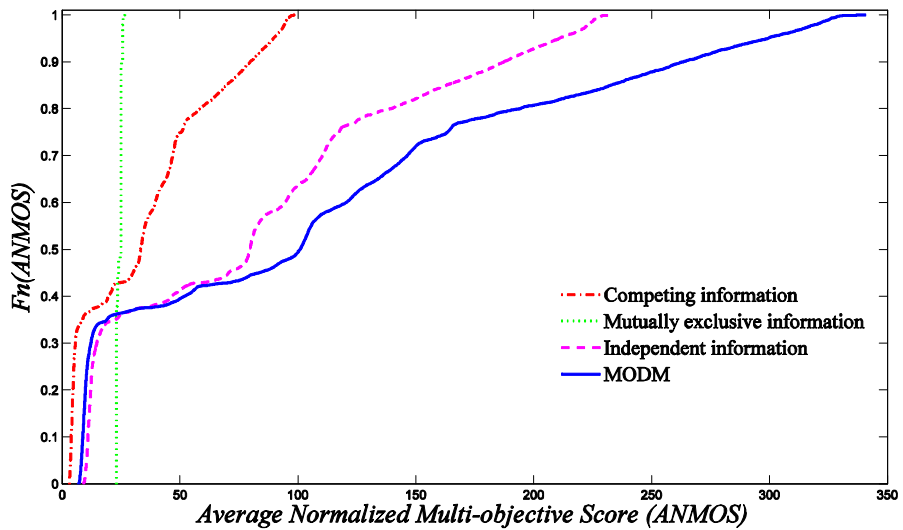


Fig. 3b Final ANMO score for MODM and single-information types

5.3.2 Comparison of MODM and single-evaluation criteria

In this experiment, we estimate the importance of MODM in comparison of single-evaluation criteria. We propagated multiple types of information in the network, but calculated their diffusion values based on single-evaluation criteria. In Fig. 4a, the initial ANMO scores are shown for all evaluation measures; again, the diffusion curves

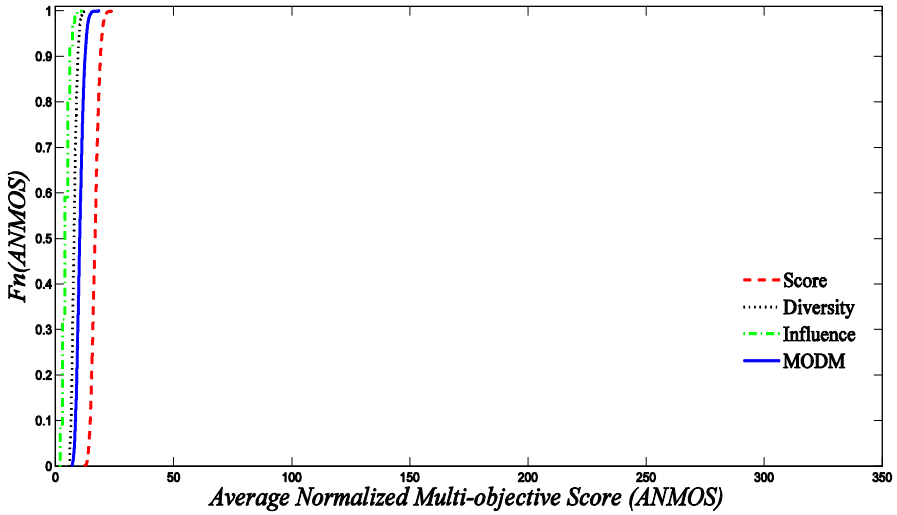


Fig. 4a Initial ANMO score for MODM and single-evaluation criteria

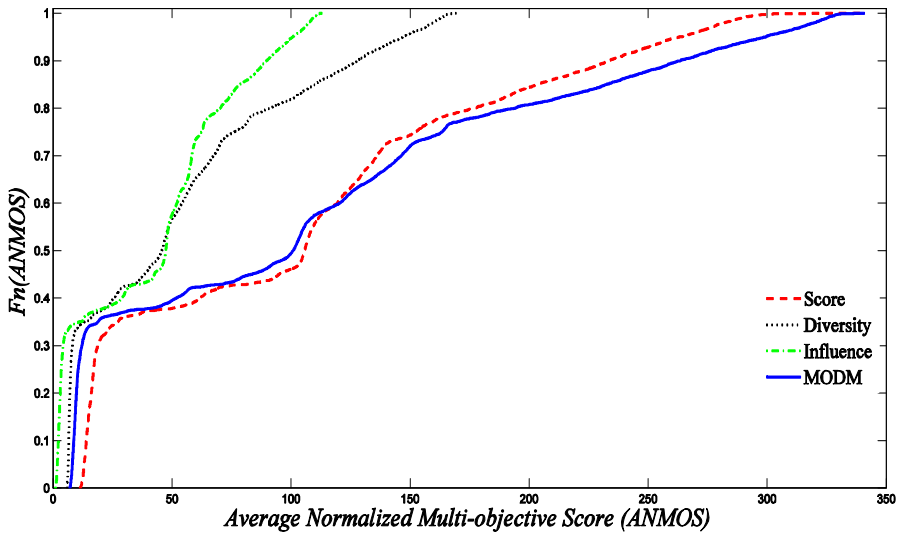


Fig. 4b Final ANMO score for MODM and single-evaluation criteria

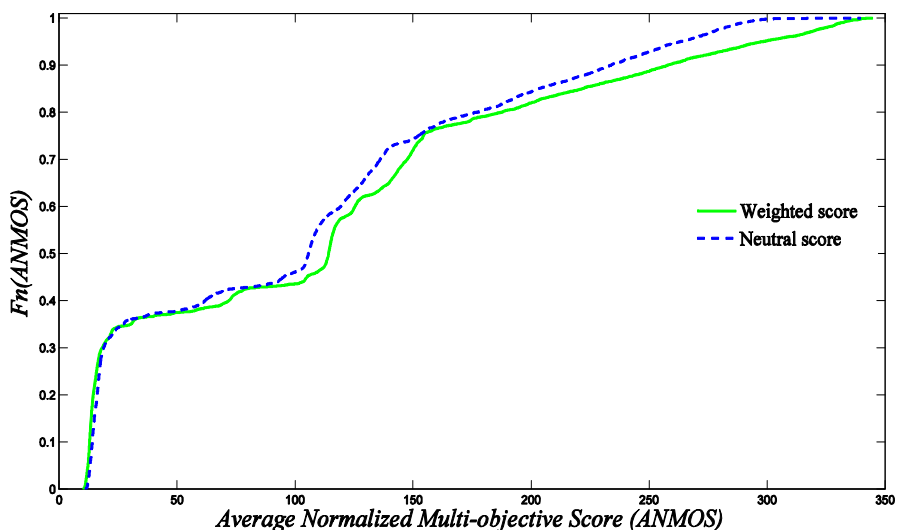
are very similar to each other. This shows that no matter whether the experiment is performed on information propagation or evaluation criterion, the initial ANMO curves are highly clustered, and there is no information dispersion in the network. In Fig. 4b, the diffusion curves in terms of the final ANMO score are illustrated for all measures. This signifies the MODM capability to propagate the information on single-evaluation criteria in addition to its original goal of achieving the multi-objectives during the diffusion process. The influence measure has very low disper-



sion; it means that the influence of the information is not changing rapidly during the interaction of individuals. Diversity measure is better than influence, as information history log helps in calculation of the diversity of each individual after each interaction. The diffusion curve for score measure gets clustered after a certain limit; however, it shows better performance than the influence and diversity measure. The diffusion curve for MODM represents its similar start with other evaluation criteria; however, it achieves the highest *ANMO* score at the end of the diffusion process that represents its significance in getting the objective of information maximization with high dispersion.

### 5.3.3 Comparison of evaluation criteria with weighted and neutral weight factors

This experiment shows the effects of weight factor attached to each evaluation criteria in Eq. (4) of MODM. In order to give high importance to particular evaluation criterion, its weight factor can be increased at the start of diffusion process. In this experiment, first, multiple types of information are propagated by assigning the equivalent weights (a neutral weight factor) to each evaluation criterion and then a higher weight of 0.6 is assigned to each evaluation criterion while the remaining weight of 0.4 is equally distributed in other measures. Diffusion curves for weighted and neutral weight factor are shown in Figs. 5a, 5b and 5c for score, influence and diversity measures, respectively. The high diffusion curves for each evaluation criterion illustrate that giving a high weight to a particular evaluation criterion thereby created a high importance in multi-objective optimization. The results of this experiment show that the proposed MODM gives flexibility to users to assign a high importance of any evaluation criterion in order to mold the diffusion process in the intended dimension for the analysis of the underlying social network.



**Fig. 5a** Score comparison for weighted and neutral weight factor

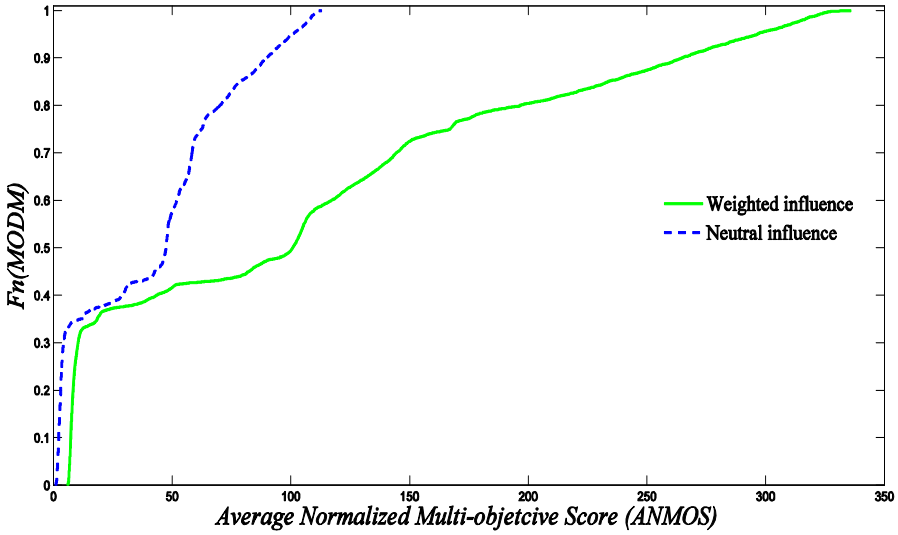


Fig. 5b Influence comparison for weighted and neutral weight factor

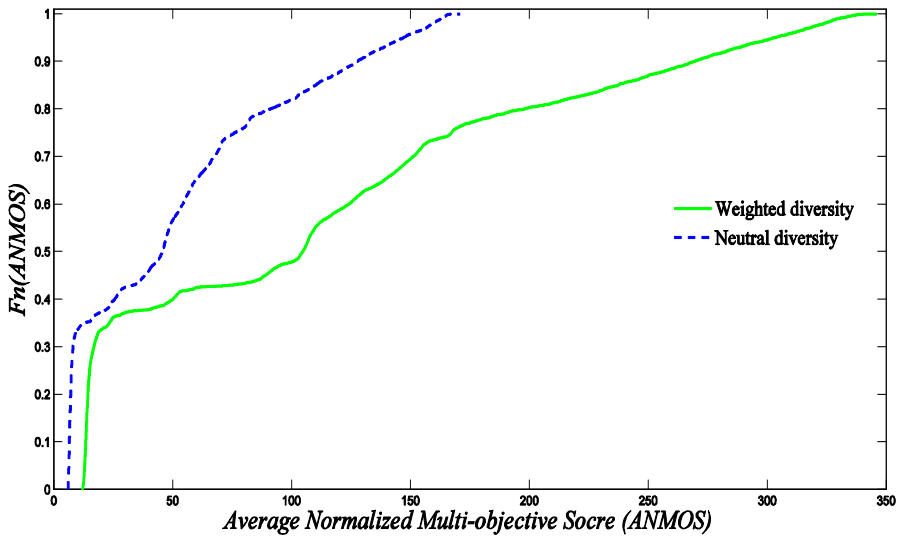


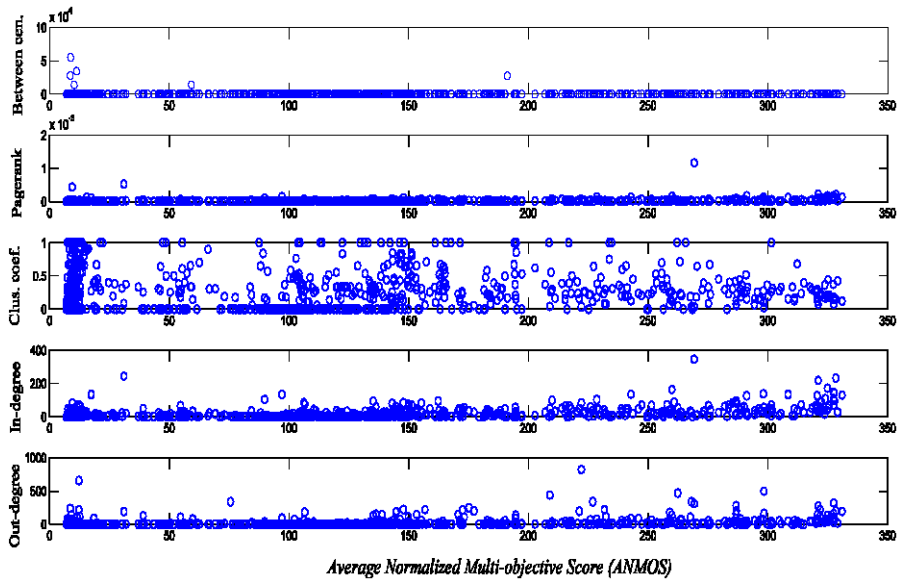
Fig. 5c Diversity comparison for weighted and neutral weight factor

5.3.4 Comparison of MODM and conventional network measures

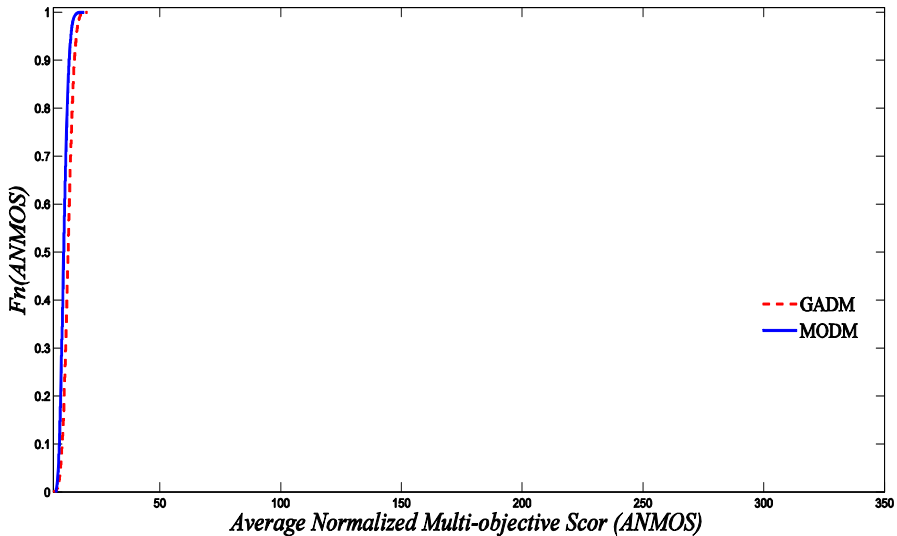
In this experiment, we compare the ANMO score of MODM with five conventional network measures: (a) Betweenness centrality: it is the centrality measure for each vertex of network that quantifies the control of an individual on the communication with others in a social network [23]. Individuals, who have a high probability to oc-

**Table 2** Correlation Comparison of *ANMO* score with network measures

Network measures	<i>ANMO</i> score	Out-degree	In-degree	Clustering coef.	PageRank
Betweenness centrality	-0.006	0.0073	-0.0143	-0.0255	-0.0051
PageRank	0.188	0.4242	0.8206	-0.0588	-
Clustering coefficient	0.146	-0.1136	-0.0387	-	-
In-degree	0.418	0.6401	-	-	-
Out-degree	0.294	-	-	-	-

**Fig. 6** *ANMO* correlation with conventional network measures

cur on a randomly chosen shortest path between two randomly chosen individuals for interaction, have high betweenness. (b) PageRank: it is a link analysis algorithm that assigns a rank to each individual within the social network based on its importance in communication [24]. If an individual has a lot of links with other high degree nodes then he is assigned high ranking. (c) Clustering coefficient: it is a measure of the degree to which vertices in a social network tend to cluster together [25]. It quantifies position of an individual in how close his neighbors are to form a complete community. (d) In-degree shows the number of incoming edges, and (e) out-degree represents the number of outgoing edges for communication. It is shown in Table 2 that high *ANMO* score cannot be explained by conventional social network measures. The correlations between *ANMO* score and other network measures are very poor as shown in Fig. 6. Intuition might suggest that an individual who receives e-mail from many people (an individual with high in-degree) would be an accumulator of information with corresponding high *ANMO* score, but the correlation between *ANMO* and in-degree is 0.41 which shows a weak correlation. Some of the measures show



**Fig. 7a** Initial Score of MODM and GADM

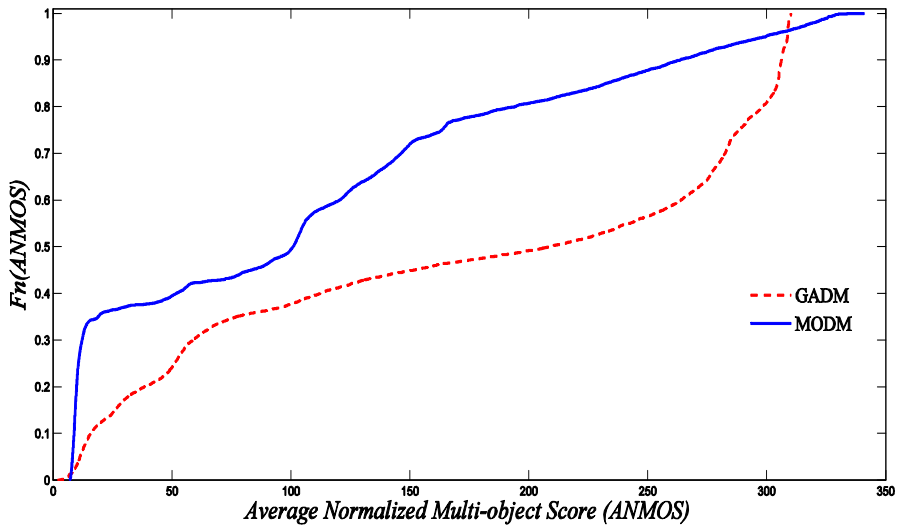
high correlation with each other, for example the correlation between in-degree and out-degree is 0.6401. The highest correlation of 0.8206 exists between PageRank and in-degree, however correlation between out-degree and PageRank is comparatively low. This shows that *ANMO* score and most of the conventional network measures are weakly correlated with each other so the high *ANMO* score in diffusion process is not dependent on the conventional measures of network.

### 5.3.5 Comparison of MODM and GADM

This experiment shows the effectiveness of MODM in comparison of GADM [18], which used an evolutionary algorithm to diffuse the information in the network as a single-objective optimization problem. We kept the settings unchanged during the experiments. In Fig. 7a, the diffusion curves for the initial score are presented, the internal processing of both the models are different. Although upon beginning the diffusion process both models show a similar state, initial scores are highly clustered and there is no dispersion. Figure 7b shows the final score curves: the diffusion curve of MODM started at the similar position of GADM, however MODM finished with high information maximization in more dispersion as compared to GADM diffusion rate. It shows that MODM outperforms the GADM in achieving a high diffusion rate in terms of an information maximization objective.

## 6 Conclusion and future work

In this paper, we proposed a multi-objective diffusion model that propagates multiple pieces of information with evolution fitness criteria by designing an evolutionary algorithm. In order to propagate multiple types of information in one diffusion process,



**Fig. 7b** Final Score of MODM and GADM

we model the set of information into a binary schema where each schema represents one type of information with its associated score. Furthermore, we maintain information history log for each individual to keep track of all incoming and outgoing information in all timestamps. This helps to predict a more accurate class of information diffusion by holding the monotonicity property about information. The information value of each individual is calculated based on evolution fitness criteria for each information type. Evolution fitness criteria utilize the benefits of score generated by the schema and information history maintained in the information history log. Our experimental results on a real-world data set show that our model is able to simulate the rich class of diffusion model and predict the information flow in the multi-objective environment. Finally, the results show that a few individuals in the network always obtain a high information rank irrespective of the start of the diffusion process.

In the future, we intend to enhance the MODM with a more realistic class of diffusion model to better understand the dynamics of diffusion process based on the underlying network. We shall investigate the possible use of genetic programming to learn about a diffusion model that matches an observed spread.

**Acknowledgements** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012047478).

## References

1. Agrawal D, Budak C, Abbadi AE (2011) Information diffusion in social networks: observing and influencing societal interests. In: Proceedings of the 37th international conference on very large data bases, pp 1512–1513
2. Fox S (2011) The social life of health information. Technical report, Pew Internet & American Life Project

3. Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the 27th international conference on human factors in computing systems, pp 211–220
4. Rodriguez M, Leskovec J, Krause A (2010) Inferring networks of diffusion and influence. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1019–1028
5. Aral S, Walker D (2011) Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Manag Sci* 57(9):1623–1639
6. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: Proceeding of the 21st ACM conference on the world wide web, pp 519–528
7. Bakshy E, Karrer B, Adamic L (2009) Social influence and the diffusion of user-created content. In: Proceedings of the 10th ACM conference on electronic commerce, pp 325–334
8. Anagnostopoulos A, Kumar R, Mahdian M (2008) Influence and correlation in social networks. In: Proceedings of the 14th internal conference on knowledge discover and data mining, pp 7–15
9. Adar E, Adamic E, Lada A (2005) Tracking information epidemics in blogspace. In: Proceedings of IEEE/WIC/ACM international conference on web intelligence, pp 207–214
10. Huckfeldt R, Sprague J (1991) Discussant effects on vote choice: intimacy, structure and interdependence. *J Polit* 53(1):122–158
11. Rogers EM (1995) Diffusion of innovations. Free Press, New York
12. Aral S, Brynjolfssen E, Alstynne MV (2007) Productivity effects of information diffusion in networks. MIT Center for Digital Business, paper 234
13. Christakis N, Fowler J (2007) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357:370–379
14. Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence in a social network. In: Proceedings of the 9th international conference on knowledge discovery and data mining, pp 137–146
15. Kempe D, Kleinberg J, Tardos É (2005) Influential nodes in a diffusion model for social networks. In: Proceedings of the 32nd international conference on automata, languages and programming, pp 1127–1138
16. Saito K, Kimura M, Ohara K, Motoda H (2009) Learning continuous-time information diffusion model for social behavioral data analysis. In: Proceedings of the advances in machine learning, pp 322–337
17. Goldenberg J, Libai B, Muller E (2001) Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*
18. Lahiri M, Cebrian M (2010) The genetic algorithm as a general diffusion model for social networks. In: Proceedings of the 24th AAAI conference on artificial intelligence, pp 494–499
19. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman, Reading
20. Payne J, Eppstein M (2006) Emergent mating topologies in spatially structured genetic algorithms. In: Proceedings of the 8th GECCO, pp 207–214
21. Holland H (2000) Building blocks, cohort genetic algorithms, and hyperplane-defined functions. *Evol Comput* 8(4):373–391
22. <http://www.cs.cmu.edu/~enron/> [Aug 25, 2012]
23. Bader DA, Kintali S, Madduri K, Mihail M (2007) Approximating betweenness centrality. In: Proceedings of the 5th workshop on algorithms and models for the web-graph, pp 134–137
24. Jing Y, Baluja S (2008) PageRank for product image search. In: Proceedings of WWW, pp 307–316
25. Liu Z, Wang C, Zou Q, Wang H (2010) Clustering coefficient queries on massive dynamic social networks. In: Proceedings of the 11th international conference on web-age information management, pp 115–126