



# A fuzzy logic approach to influence maximization in social networks

Yacine Atif<sup>1</sup> · Kanna Al-Falahi<sup>2</sup> · Tshering Wangchuk<sup>3</sup> · Birgitta Lindström<sup>1</sup>

Received: 3 January 2019 / Accepted: 26 March 2019 / Published online: 12 April 2019  
© The Author(s) 2019

## Abstract

Within a community, social relationships are paramount to profile individuals' conduct. For instance, an individual within a social network might be compelled to embrace a behaviour that his/her companion has recently adopted. Such social attitude is labelled social influence, which assesses the extent by which an individual's social neighbourhood adopt that individual's behaviour. We suggest an original approach to influence maximization using a fuzzy-logic based model, which combines influence-weights associated with historical logs of the social network users, and their favourable location in the network. Our approach uses a two-phases process to maximise influence diffusion. First, we harness the complexity of the problem by partitioning the network into significantly-enriched community-structures, which we then use as modules to locate the most influential nodes across the entire network. These *key users* are determined relatively to a fuzzy-logic based technique that identifies the most influential users, out of which the seed-set candidates to diffuse a behaviour or an innovation are extracted following the allocated budget for the influence campaign. This way to deal with influence propagation in social networks, is different from previous models, which do not compare structural and behavioural attributes among members of the network. The performance results show the validity of the proposed partitioning-approach of a social network into communities, and its contribution to “activate” a higher number of nodes overall. Our experimental study involves both empirical and real contemporary social-networks, whereby a smaller seed set of *key users*, is shown to scale influence to the high-end compared to some renowned techniques, which employ a larger seed set of *key users* and yet they influence less nodes in the social network.

**Keywords** Social networks · Community detection · Influence propagation · Fuzzy logic

## 1 Introduction

The impact of online social networks (OSNs) has undeniably affected a sizeable proportion of the world population who in way or the other, tend to use YouTube, Facebook, Twitter, Flickr, MySpace and LinkedIn, etc. The impact of social networks on individuals' behaviour throughout various

stages of their life has been extensive, and in different circumstances. Subsequently, social networks have become a prime venue for propagating influences using different techniques, and disseminating information of all kinds. This phenomenon is facilitated by social connections which spread information from one individual to another at a faster pace, particularly when critical events arise. For example, tweets (i.e. Twitter posts) have considerably increased in volume during the severe 2011 Tsunami in Japan (Acar and Muraki 2011), during which individuals around the devastated areas posted tweets to alert followers about their situation. Similarly, and in the same year, the political unrest in Egypt and Tunisia were driven by bloggers posting their exasperation against their respective government practices, on social networks. The extraordinary observation during these events, is that massive physical-protestations took place on streets following virtual frustration expressions to get rid of dictatorships. This illustrates the influential power of social networks, whereby actions are not just embraced online,

---

✉ Yacine Atif  
Yacine.Atif@his.se

Kanna Al-Falahi  
k.alfalahi@uaeu.ac.ae

Tshering Wangchuk  
tshering\_wangchuk@rim.edu.bt

<sup>1</sup> School of Informatics, University of Skövde, Skövde, Sweden

<sup>2</sup> College of Information Technology, United Arab Emirates University, Al-Ain, UAE

<sup>3</sup> Royal Institute of Management, Thimphu, Bhutan

but individuals do engage into translating them across the physical world also.

Social networks have long been a nature of the humanity life, where distinguished individuals of a community could drive members of their community into embracing a faith or adopting a behaviour or change their life conditions (Khoua and Atif 2018). This natural humanity trait, has subsequently been digitalised in OSNs. However the propagation pace in OSNs is much faster than in real-life networks, and connections between OSN users scale much higher and quicker in OSNs leading to increasingly new relationships and new community memberships. Like in real-life social circles, people expand and benefit from their social relationships in OSNs as well. Traditionally, influencers tend to be those who accumulate many relationships. However, social ties develop also when the same action cascades over OSNs, which generates a propagation wave that may reach many more individuals than direct connections. This is how bloggers “like” a particular topic, and spread the induced message that spawns growing discussions, across contemporary social networks (Dumenco 2011). Similarly, a product or a service may be embraced by many individuals across intermediate connections, as well. Subsequently, the concept of interest graphs evolved, where nodes designating individuals express their mutual interests for a content node (Solis 2011). Hence, there are two types of node, and two types of links as well, that link people to content of their interests, and content to content to express content relationships. This graph concept supports brand evolution by intersecting interest graphs to build larger communities that are used to spread influence in a targeted advertisement campaign for example. This approach has later evolved further into a major marketing trend in contemporary OSNs (Solis 2011).

OSNs are investigated using concepts from graph theory (Newth 2006) in the context of computer science and social science (Scott 2000) disciplines, leading to an interdisciplinary social network analysis (SNA) area. Therefore, an interdisciplinary approach is required to investigate and evaluate OSNs. A graph representation of OSNs include nodes or vertices that represent OSN individuals, and links or edges that represent social ties, such as friendship. This representation is frequently used in the evolving SNA field (Hanneman and Riddle 2005) to identify the triggers and distribution patterns of influence propagation in social networks. The influential power of a user augments with the relationships he or she develops with peers who have varying degrees of influence power, in the social network. SNA attributes of a node centrality such as degree, closeness and betweenness centralities (Hanneman and Riddle 2005) were used earlier to quantify nodes’ power in the network. The influence power of a node grows with the node’s degree, closeness or betweenness values. A graph-theory structure of a social network sample is depicted in Fig. 1. The graph representation

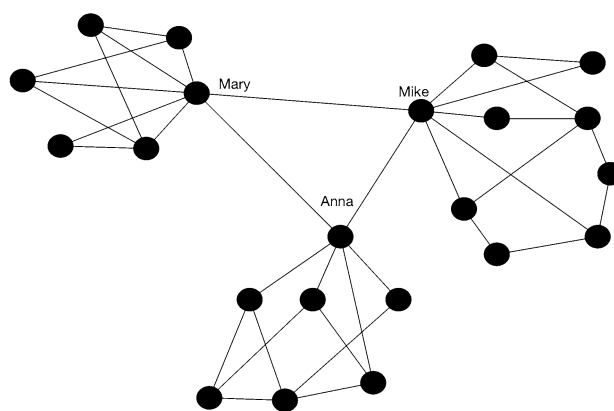


Fig. 1 Social network diagram (Source: <http://www.fmsasg.co/SocialNetworkAnalysis>)

displays groups of nodes with some central ones that can trigger a great influential campaign when disseminating their social influence. However, centrality attributes alone reflect only structural aspects of a node. We argue in this paper that the influence assets of a node are also driven by the dynamic impacts across links and connections originating from that (influencer) node within OSN. This impact occurs when users along connecting paths to the influencer embrace the influencer’s advocated behaviour or embrace the influencer’s action, labelled throughout this paper as “common-action”.

We propose an approach by which a set of  $k$  nodes in a social network are discovered based on both their structural and historical-actions attributes to maximize influence propagation. The proposed approach employs three successive processes that work in tandem. Initially, “artificial” communities are built whereby “similar” users are assembled together, based on a judicious similarity function. Next, for each of these synthetic communities, we identify *key users* using a computational intelligence technique that employs a fuzzy-logic based function to discriminate nodes based on both their structural-centrality and influence-weight attributes. To measure influence weights, we suggest to crawl action-logs across nodes to figure out instances of common-actions adoption. Obviously, these measurements are subject to dynamic changes, depending on the accumulated behaviours of users from past activity logs in the social network. Finally, we rank these *key users* based on their influence power, by simulating an influence diffusion process to determine the *seed-set* of candidate influence propagation nodes. The rationale to pre-process the original social network through the identification of virtual communities, is inspired from the fact that members of the same community tend to think the same and hence they facilitate the propagation of incoming influences from peer members of their own community. The final step in the above three-steps process predicts the most influential members from candidate *key*

*nodes* based on the available marketing budget to shortlist agents for promoting products or services and recommending them around social networks.

The rest of this paper is organized as follows: Sect. 2 provides a background and related works about the main areas that are relevant to our proposed work. Section 3 depicts our proposed community-aware social influence diffusion approach. Section 4 demonstrates the efficiency of our proposed approach through an experimental analysis, based on data sets from simulated and existing social networks. We wrap-up the suggested developments discussed in this paper within Sect. 5, where we also reveal some of our ongoing extensions.

## 2 Problem and background

Members of a social network are expected to build connections with other members of the network. To model members and inherent connections, a graph  $G(N, E)$ , represents the set of members  $N = \{1, 2, \dots, n\}$  which is implemented as an  $n \times n$  adjacency matrix. The influence weights:  $0 \leq E_{ij} \leq 1$  are the adjacency matrix entries. Thereby the graph  $G(N, E)$  is said to be a weighted graph. When  $E_{ij} \neq E_{ji}$ , the graph is said to be directed and undirected otherwise.

An influence occurs when a user  $u$  of a social network represented by the graph  $G$  embraces a behaviour, that was previously embraced by another user  $v$ , in which case  $v$  is said to have *activated*  $u$  or  $u$  is said to have been activated. Other nodes of  $G$  that do not perform the action or embrace the behaviour of  $v$  are said to remain inactive. Subsequently, influence maximization consists in discovering a subset of *key* users  $U$  in the social network modelled by the graph  $G$ , where  $|U| = k$ , who activate as many users in the social network as possible. Identifying  $U$  is the core problem of influence maximization (Goyal et al. 2010). The discovery of key users is subject to a multi-criteria decision-making dilemma due to the contribution of both nodes' influence weights and their topological attributes within  $G$ . This dilemma motivates the rationale of our proposed computational intelligence technique based on a fuzzy-logic model to maximize influence, which forms the main contribution of this paper. To understand further this rationale, we introduce some relevant fuzzy-logic concepts and use them to illustrate our proposed approach.

### 2.1 Fuzzy logic

Fuzzy logic is a prominent development in computational intelligence (Zadeh 1965). This theory tolerates logical assertions to carry a progressive extent of values that lie within the interval  $[0, 1]$  as an alternative to true/false assertions (Hellmann 2001). The approximation in the reasoning

processing led to several applications in contrast to its crisp counterpart, and appears more natural to mimic human, rather than machine reasoning (Zadeh 1984) to evaluate real-world considerations. We adopted fuzzy-logic to break the dilemma induced when selecting highly influential nodes, thereafter labelled "key nodes", out of which we pick the seed-set of nodes to use for an actual influential propagation instance, in order to meet marketing budget limitations. The joint criteria used to determine "key nodes" membership is found to be effectively addressed using a fuzzy-logic based membership function.

Fuzzy-logic sets are characterised with partial membership features, unlike crisp set counterparts (i.e. either an element is a member of set or not), and thus they adapt better to natural membership expressions used in real-world situations (Baig et al. 2013). A membership function is used to evaluate the extent of membership and which is context dependent to meet the realistic real-world features (Rahman and Ratrouf 2009). The membership function computes the actual membership extent within the interval  $[0, 1]$  to assert a statement with a certain context-related degree, that contrasts with traditional logical assertions with exclusively true or false propositions (Rojas 1996).

### 2.2 Illustrative scenario

To illustrate the application of fuzzy-logic to our proposed approach to select "key nodes", consider the following example that is adapted from (Wolfram 2014). Consider the problem of discovering the most influential users in a social network made up of the following *nodes* =  $\{1, 2, 3, 4, 5\}$ . The goal is to determine nodes that combine centrality location and influence-weight on other nodes attributes. A first fuzzy set is developed to represent the centrality degree as follows:

$$\text{Centrality}(C) = \{\{1, 0.4\}, \{2, 0.6\}, \{3, 0.8\}, \{4, 0.4\}, \{5, 0.5\}\}.$$

Note that Node 3 is deemed most central given its membership value or grade. However, Nodes 1 and 4 are the least central nodes in the network. The other fuzzy set *Influence Weight* ( $IW$ ) =  $\{\{1, 0.1\}, \{2, 0.9\}, \{3, 0.7\}, \{4, 1\}, \{5, 0.2\}\}$  represents the influence power of each node. Note that the influence weight here represents an averaged value of influence that a node has over all nodes in the network. A high value indicates a high influence power. Therefore, Node 4 is deemed to be the most influential user in the entire network.

Next, we engage into a decision-making process to identify the set of nodes that are most influential considering both of the above constraints, i.e. favourable location vs. influence weight. The natural answer to these joint criteria is a fuzzy intersection of both membership sets to identify nodes that optimise both features, simultaneously. A fuzzy intersection considers members with

lowest grades in each of the fuzzy sets *Centrality* ( $C$ ) and *InfluenceWeight* ( $IW$ ) (Rojas 1996), thus resulting in  $C \cap IW = \{\{1, 0.1\}, \{2, 0.6\}, \{3, 0.7\}, \{4, 0.4\}, \{5, 0.2\}\}$ . This decision-making process is illustrated in Fig. 2. The maximum grade value grants highest influence feature to a user. It appears that Node 3 is elected as the one with the highest grade and thus is deemed to be the most influential in the network considering both constraints.

There were fuzzy-logic considerations to model social networks. However, they were limited to studying common network attributes such as degree, clustering and betweenness (Kundu and Pal 2015). This approach focused on fuzzy relationships among social network users. Similarly, distinguished relationships among actors in social networks were modelled using “fuzzy graphs” in Nair and Sarasamma (2007). Our approach advocates fuzzy logic to select key nodes in our community-driven influence propagation approach.

### 3 Related works

The process of identifying key users in a large network such as those found in contemporary social networks, can be harnessed by decomposing the network into communities. Intuitively, the influence propagation process is expected to spread faster among members of the same community with shared interests. Subsequently, community detection is combined with influence maximization in this paper, and thus our review of existing works encompasses both areas.

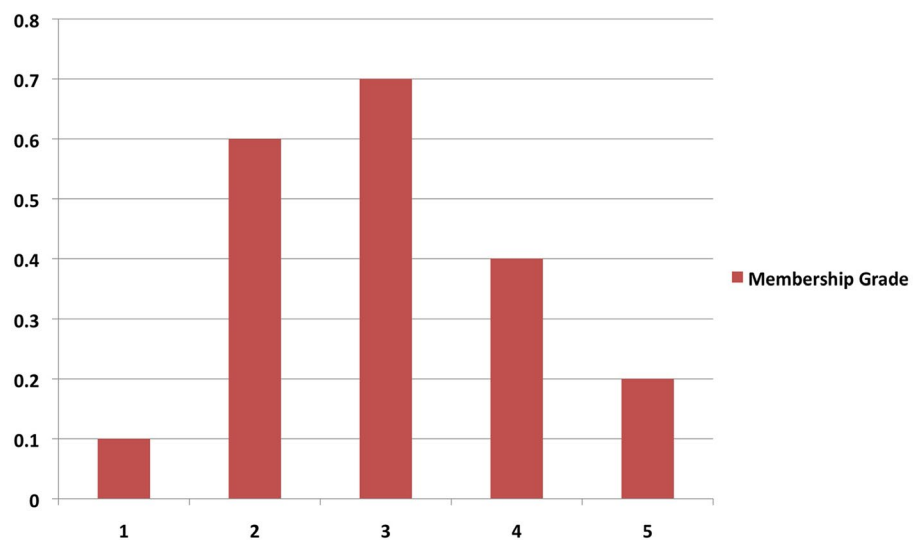
#### 3.1 Community detection

Identifying members within their circle of common interests has been a vector to direct marketing campaigns according

to the interests of the social circle members. However, this identification process requires the discovery of social network members with shared interests. One of the prominent works found in the literature discovers these clusters of social network members by hierarchically dividing the network through eliminating iteratively network edges (Newman and Girvan 2004). This process leads to a division of the network into dense clusters of users, thereby leading to community structures. The candidate edges for removal are those with high-betweenness value. This value that is associated to a candidate edge, quantifies the length of the shortest-path between any two nodes, when that path passes through the candidate edge. However, as edges are taken out from the network, all betweenness-values need to be recomputed since the paths based on which the previous computation was made may have changed. A desired threshold is used to evaluate the quality of the detected communities in each iteration to decide whether to stop the network division process. Nevertheless, this technique is seldom employed for identifying communities due to its complexity and incurred computational costs. Instead, the opposite agglomerative alternative is mostly used. A hierarchical clustering approach built upon the above technique discovers and takes out edges iteratively from the network based on a centrality value (Fortunato et al. 2004). The authors show the effectiveness of their approach, despite the  $O(n^4)$  complexity of the proposed algorithm.

A quality-driven division approach has also been proposed using a metric called modularity (Newman 2004). Labelled  $Q$ , the modularity is a function that sizes the significance degree of detected communities. This approach is distinguished by its simplicity and viable worst-case computational complexity of  $O(n^2)$ . Subsequently, this approach has been deemed attractive and employed in several applications. Nevertheless, the modularity is upper-bounded by a

Fig. 2 Fuzzy decision plot



threshold, which is a function of the network cardinality, and communities with modularity values lower than that threshold could not be detected. To overcome these bounds, a metric that measures communities' density has been suggested Li et al. (2008). This approach employs both vertices and edges, while preserving the iterative division process of the network to detect communities. Yet, this proposed procedure is NP-hard. An alternative resolution of the upper-bounds problem involves a variation in the modularity function formula (Arenas et al. 2008). Later, it was shown that such variation of the modularity function exhibits also boundary issues when combining smaller clusters and dividing larger ones (Lancichinetti and Fortunato 2011).

As mentioned earlier, an alternative approach to divisive techniques was the agglomerative one (Clauset et al. 2004a) suggested by the same authors, and labelled following their initials CNM. This method proceeds from the bottom of the dendrogram that hierarchically displays the relationships between nodes, and move up in a greedy way, while assembling clusters of the network. Although analogous to (Newman 2004), this approach has better complexity performance of  $O(n \log_2 n)$  in worst-cases.

### 3.2 Influence maximization

The influence maximization problem has been extensively investigated in the literature, particularly in quantifying the expectation of a node to influence other nodes. But existing approaches face a limited capacity to maximize the activated social network nodes to the higher end, while minimizing the seed-set size  $k$  of selected nodes used to propagate influence. Constant probabilistic values are assigned to nodes in the static approaches to model influence propagation within social networks based on time-independent observations (Goyal et al. 2010). However, these methods assume a static propagation of influence that do not evolve over time, given the constant probabilistic values. This means, they do not address the development of influence probabilities following users' activities in the social network. The Bernoulli probability distribution was used in the above static modeling approaches to represent social network users attempting to activate neighbouring peers. The influence propagation models using static approaches are simple to use, but the natural evolution of social networks limits their applicability. The induced constant probabilities assumption oversimplifies influence measurements to accommodate contemporary social networks.

Dynamic approaches to represent influence propagation such as the Snapshot approach (Kossinets and Watts 2006; Backstrom et al. 2006; Shi et al. 2009) does consider the evolution of probabilistic values to reflect the nodes' evolving influence power over time. As the name implies, this approach considers successive snapshots of the network over

time to infer its evolution. This approach has been extensively used given its capacity to pick up the dynamics of social network data for analytical purposes, including the evaluation of influence state among nodes across successive timestamps. However, consecutive snapshots increase substantially the size of data to analyse. Alternatively, ordinal-time approaches limit the observation sequences to activation occurrence instants (Cosley et al. 2010). That is, when there is a change in the network induced by an influence-related activity, a snapshot of the network is retrieved, which lowers the size of data to analyse. Nevertheless, timestamped snapshots of an entire social web structure are complex to collect, which reduces the implementation efficiency of related approaches, that aim at evaluating activation patterns across influence-propagation processes.

Alternative approaches to model influence that appear to be less sensitive to the above drawbacks have been proposed. The landmark Linear Threshold Model (or LTM) and Independent Cascade Model (thereafter labelled ICM or IC) fall in this category. LTM (Domingos and Richardson 2001; Kempe et al. 2003; Richardson and Domingos 2002) accumulates the influence weight contribution from each node towards a common neighbour. When the resulting accumulated value exceeds a threshold, the common neighbour is activated. Edge weights reflect the influence power a node may have over his neighbours. ICM (Kempe et al. 2003), advocates a binary states of nodes whereby each node has a single chance to be activated or not during an influence-diffusion that cascades over neighbouring nodes. Activated nodes will have the same chance to activate their neighbours, recursively. This process is similar to viral spreading across ties in conventional social networks where users incite peers to watch the same movie, or embrace a certain political opinion. Subsequently, a cascade is enacted which diffuses the influence over the network structure. Activation occurs at a given node based on some probabilistic value, which evolves according to the interaction intensity between nodes. These approaches speed up the influence propagation process, particularly when the seed-set of highly influential users is pre-established.

However, the above LTM as well as IC models do not consider the mutual relationships among node actions. This observation called for alternative approaches that consider users' actions towards a common context. Topical graphs mine users' activities using a machine learning approach to infer influence probabilities following users' interest in particular topics (Tang et al. 2009). A related subsequent investigation found that similar users tend to influence each other (Sun and Tang 2011). This relationship between similarity based on social ties and influence activation supports further our influence-propagation approach and our rationale for our proposed community-driven influence propagation. However, the effectiveness of influence propagation is enhanced

by decreasing the seed set of highly-influential nodes (Hosseini-Pozveh et al. 2017) and harness the complexity by using a modular propagation approach, such as those that employ communities (Wang et al. 2010), like we do. But these approaches are developed for specific purposes and do not incorporate computational intelligence techniques to optimise the seed-set selection as we do with our proposed fuzzy-logic based influence propagation model.

#### 4 Community based influence propagation algorithm

In this section, we reveal our approach to influence-maximization which includes a community-enrichment preprocessing step to scale-up the diffusion process and the number of activated users within a social network. In doing so, we join together two of our previous works, namely an original approach to identify communities (AlFalahi et al. 2013) and a technique to evaluate influence weights (AlFalahi et al. 2014). The combination of these works generates a new approach whereby the previous techniques are employed in tandem, to obtain a set of users who can incite neighbouring peers to embrace an advocated behaviour. The analysis results shown later in the experiments section, reveal the effectiveness of this new approach to maximize influence on synthetic and existing social networks data.

Our suggested approach consists actually, of three successive steps, namely (1) Communities identification via an enhanced Similarity-CNM network algorithm, (2) Key Users discovery within each detected community, and (3) Seed set identification from ranking key users, to effectively drive the diffusion process over neighbouring peers. Figure 3 illustrates our proposed framework and Table 1 provides an explanatory reference of the symbols used throughout this paper.

##### 4.1 Similarity-CNM

Given an input social network, our proposed approach starts by discovering communities. This essential step of our approach employs a similarity function to support behavioural embracement among similar peers. This preprocessing step ensures that the search space for *key users* is reduced into modular communities and facilitates further the subsequent diffusion process. Our inspiration that is supported also by previous investigations (Wang et al. 2010), is that users with high similarity-attributes are more susceptible to embrace common attitudes. Thus, the community structures which first assemble similar users into modular communities facilitate the process of key-users discovery. These key users are the first seed-set candidates to propagate influence, that are further ranked to extract a subset that meets

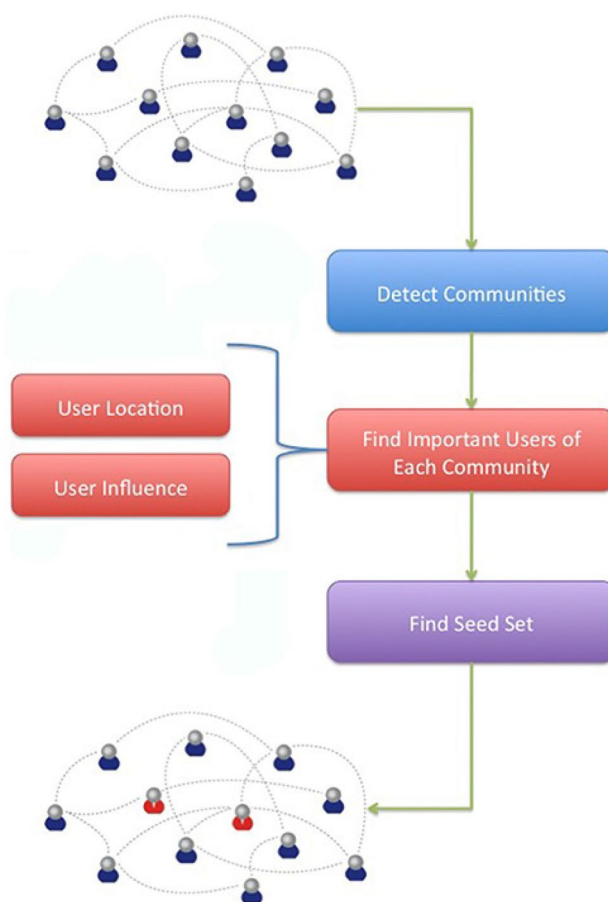


Fig. 3 Proposed framework

some budgeting resources allocated to a given marketing campaign.

An improved version of the CNM algorithm (AlFalahi et al. 2013) is shown in Algorithm 1 depicted next. Named Similarity-CNM, this approach detects communities through an improved version of the existing CNM landmark approach (Clauset et al. 2004a). Based on the performance results revealed later in this paper, the quality of the communities from the improved CNM-Similarity version outperforms the original CNM. Subsequently, the influence modelling steps follows the community detection one, using a network of modular virtual-communities instead of the original plain network. This community-enriched network is deemed to supply additional information that guide further the spread of influence across the entire network. The discovery of communities is preceded by enriching the network with synthetic links that join similar nodes together, in order to obtain denser community structures. The preprocessing step incurs a computational complexity of  $O(n^2)$ . However, this preprocessing step is carried out offline to alleviate this additional computational cost throughout the influence maximisation process.

**Table 1** Legend of symbols

Symbol	Legend
$G$	Social network graph
$N$	Nodes set in a social network
$n$	The number of nodes in the network, i.e. $ M $
$E$	Edges set in a social network
$D_i$	Degree of Node $i$ , i.e. the number of ties that Node $i$ has in the network
$E_{ij}$	Adjacency matrix entry corresponding to Nodes $i$ and $j$ , i.e. = 1 if there is a link between $i$ and $j$ , and 0, otherwise.
$n_i$	Number of nodes adjacent to Node $i$ , i.e. number of $i$ 's neighbours.
$cn_{ij}$	Number of common adjacent nodes to Nodes $i$ and $j$
$C$	A community in the social network
$Q$	Modularity value of a social network with respect to community structures, and determined by Eq. (7)
<i>Similarity_Threshold</i>	Parameter used to determine similar nodes with respect to the value delivered by Eq. (1)
<i>CentralityWeight<sub>i</sub></i>	Level of Node $i$ centrality in the network, and determined by Eq. (2)
<i>Threshold_IU</i>	Parameter used to determine the number of important users in a network with respect to their centrality weight values
<i>Degree Centralities</i>	Vector of centrality weight values of nodes $i$ , $\forall i \in N$
<i>Central Users</i>	The first <i>Threshold<sub>IU</sub></i> nodes with highest centrality weight
<i>Threshold_S</i>	Parameter to set the desired size of influential-nodes seed set
<i>InfluenceWeight<sub>ij</sub></i>	Influence weight of Node $i$ over a particular Node $j$ , given $E_{ij} = 1$ and determined by Eq. (3)
<i>Influence Weights Avg<sub>i</sub></i>	Average of influence weights that Node $i$ has over all nodes $j$ , given $E_{ij} = 1$ , and determined by Eq. (4)
<i>Intersection<sub>i</sub></i>	A fuzzy intersection value, between <i>CentralityWeight<sub>i</sub></i> and <i>InfluenceWeightsAvg<sub>i</sub></i> membership grades determined by Eq. (5)
<i>Key Users</i>	A set of nodes with both favourable location and influence history determined by Eq. 6

The preprocessing step results in an identical network from the original one, with added links that connect similar users, thereafter labelled similarity-network. A directed unweighted graph representation of the network is considered, where users' similarity represent weight annotations over the network edges. These similarity weights provide indicators about users' relationships to construct well-structured communities. The discovery and evaluation of community structures from the similarity-network uses the standard CNM algorithm (Clauset et al. 2004b). As stated earlier, this algorithm clusters network nodes iteratively following an agglomerative approach that moves up through the hierarchical network structure, while joining clusters together. The modularity is computed on the way up, to evaluate the clusters' quality  $Q$ , which represents the variation of links within clusters and a presumed number of links. Good structures rise with such variations (Newman 2004). Initially a small set of nodes is built up without any links, and hence with a poor modularity. The community structure is iteratively enriched with edges while merging cluster pairs, which raises modularity values. The prospects of building sparse communities (Fortunato 2010) is reduced by supplying CNM with enriched similarity-network. We refer to this combined algorithm and enriched input as Similarity-CNM approach. The algorithmic steps to obtain the virtual similarity-network  $G'$ , given an input network  $G$ , are revealed

in Algorithm 1. The employed similarity function is shown in Eq. (1):

$$\text{Similarity}(i, j) = \frac{E_{ij} + cn_{ij}}{n_i + n_j}. \quad (1)$$

Equation (1), makes use of adjacency values  $E_{ij}$  that represent the size of common-neighbours  $cn_{ij}$  between the nodes  $i$  and  $j$ , using their respective degree  $n_i$  and  $n_j$ . The proposed pre-processing step results in more inclusive communities and speeds-up the community-detection process. The objective is to improve the structure of detected communities with high-modularity values. The discovered virtual communities are used to find candidate key users. This process begins by identifying users with highest centrality values within each community. Then, the nodes with highest influence weights are obtained.

The complexity of the algorithm is of order  $O(n^2)$ , but since the process is performed offline as a preprocessing step to virtually transform the input network, the complexity doesn't really affect the performance of the framework. The next step of the framework is to establish communities in the enriched network  $G'$  by applying some contemporary community detection algorithms, such as CNM (Clauset et al. 2004b).

---

**Algorithm 1** Similarity-CNM Algorithm
 

---

1.  $Similarity\_Threshold \leftarrow$  Threshold parameter used to determine similar nodes
  2. Copy Graph  $G(N, E)$  into the targeted similarity network  $G'(N, E')$
  3. Calculate nodes similarity for each node in  $G$  based on Equation 1
  4. For each node  $i$  and  $j$  in  $G$ :
    - (a) If  $i$  and  $j$  are similar (i.e.  $similarity(i, j) < Similarity\_Threshold$ ) then establish a virtual link between them in  $G'$ .
  5. Save the similarity enriched network  $G'$
- 

We show later in the experimental analysis section, that the community quality is highly optimized when the proposed pre-processing step of Algorithm 1 is applied. Next, we show how this virtual division of the network into well-structured communities contributes to elicit nodes used to propagate influence.

## 4.2 Key users

Key users are discovered initially from the the communities generated in the previous Similarity-CNM algorithm step. They represent seed-set candidates to propagate influence. They are distinguished by their favourable position in the network which is quantified through structural centrality values, and their influence-weight which is quantified from historical logs data. As stated earlier, to break the dilemma of dealing with dual-criteria simultaneously, we employ fuzzy-logic theory (Kahraman 2008) to select key users that optimise both criteria. In doing so, we identify the attributes involved in the key-users membership-function, as well as the associated weights to reflect the importance of some attributes over others (Peneva and Ivan 2008). The attributes here are the centrality and the influence power of users in the network, whereas the weights are importance parameters associated with each of these two criteria. Following the definition of criteria and associated weights, key

users are elicited using the fuzzy-logic process shown in Algorithm 2, which we elaborate further next.

### 4.2.1 Central users fuzzy set

The structural attribute of key users reflect their favourable position, such as the ones with high-degree values, or those bridging two or more clusters, who have the capacity of carrying influence across cluster users. These are examples of key user structural attributes, which are some interpretation of user *centrality*. We adopt Degree Centrality to measure structural attribute values. Central users fuzzy-set is determined with these values derived from the corresponding membership function, which we discuss next.

Initially, and for each detected community, a Degree Centrality is computed. To do this, the out-degree and in-degree are of each user node are determined and then cumulated. To recognise central users, we employ a *centralityThreshold*, whereby user nodes with degree exceeding *centralityThreshold*, are deemed structurally *central*, and will be carried forward to the next stage. Based on this approach, the degree centrality for all users is computed using the following membership function to determine central users:

$$CentralityWeight_i = \frac{D_i}{|E|}. \quad (2)$$



---

**Algorithm 2** Key Users Identification

---

1. Obtain Similarity-network from the input graph  $G$  (Algorithm 1: Similarity-CNM)
  2. Detect Communities  $C$  in the enhanced Similarity-network from Step 1 (Apply Community Detection algorithm)
  3. Set  $Threshold\_Important \leftarrow$  Number of Important Users
  4. Set  $Threshold\_Seed \leftarrow$  Number of Users in the Seed Set ( $S$ )
  5. Find Central Users fuzzy set
    - (a) For each detected community in  $C$  do
      - i.  $CentralUsers \leftarrow$  Determine Central Users ( $DegreeCentralities, Threshold\_Important$ )  
// Using centrality weight values and a threshold to qualify central users
      - ii. For each  $i \in CentralUsers$  do
        - A.  $CentralityWeight_i = D_i/|E|$   
/\*Membership function of Central Users fuzzy set, where  $D_i$  is Node's  $i$  degree and  $|E|$  is the total number of network edges \*/
  6. Determine the Influence Weights fuzzy set
    - (a) For each Node  $i \in CentralUsers$  do
      - i.  $InfluenceWeight_{ij} \leftarrow CalculateInfluenceWeight(i, j)$   
/\* Determine influence weights of Node  $i$ , for every Node  $j$  in the network using Jaccard Coefficient based on historic Common Actions using Equation 3\*/
      - ii. For each  $i \in CentralUsers$  do
        - A.  $InfluenceWeightsAvg_i = \frac{\sum_{j=1}^n InfluenceWeights_{ij}}{n}$  /\*Membership function of Influence Weights fuzzy set\*/
  7. Find Key Users (Fuzzy Decision)
    - (a) For each  $i \in CentralUsers$  do
      - i.  $Intersection_i = \min(CentralityWeight_i, InfluenceWeightsAvg_i)$
    - (b)  $KeyUsers \leftarrow \operatorname{argmax}_{i \in N}(Intersection_i)$   
// Select the nodes with maximum intersection of membership-grades to decide the key users
  8. Find Seed Set  $S$ 
    - (a) For each Node  $i \in KeyUser$  do
      - i.  $Reachability_i \leftarrow$  Apply Influence Propagation Method to every Node  $i \in KeyUsers$  set (Algorithm 3: Modified Independent Cascade)
    - (b)  $S \leftarrow$  Select the top influential users  $i$ , with highest  $|Reachability_i|$ , up to  $Threshold\_Seed$  members of  $KeyUsers$  set
    - (c) Return  $S$
- 

In Eq. (2),  $D_i$  represents Node's  $i$  degree, which cumulates the in-degree and out-degree of Node  $i$ . The overall number of links in the network is formulated by  $|E|$ . Structural centrality values are thus determined by Eq. (2) which define Central Users fuzzy-set. These values fall within the interval [0.1], to reflect the centrality extent of each network node. Those nodes with close to 1 centrality value, indicate a high-centrality position. Central Users fuzzy-set values are employed in the selection process of key users, as discussed next.

**4.2.2 Influence weights fuzzy set**

In addition to the structural attribute, key users are also determined based on their capacity to spread influence across the network. The computation of this influence capacity determines the influence-weight value of a user, which is formulated using a Common Actions version of Jaccard coefficient (AlFalahi et al. 2014). As an illustration of this computation, suppose a user node  $A$  triggers

a behaviour at timestamp  $T1$ , and at a later stage User  $B$  embraces that behaviour at timestamp  $T2$ . This sequence of events indicates that an activation instance occurred when User  $B$  adopts the behaviour initiated by User  $A$ . To calculate the Common Actions Jaccard coefficient, we enumerate the actions that a user adopted, and that were previously triggered by a neighbouring user in the network. The real-world experimental data we used shows that an action is triggered by a single source, and thus this consideration is assumed throughout our proposed influence-propagation algorithm. Equation (3) shows the actual formulation of the common actions Jaccard coefficient.

$$InfluenceWeight_{ij} = \frac{A_{ij}}{A_i + A_j - A_{ij}}. \tag{3}$$

In Eq. (3),  $A_i$  represents the number of actions accomplished by Node  $i$ ,  $A_j$  represents the number of actions accomplished by Node  $j$  and  $A_{ij}$  represents the number of common actions,

that represent those actions accomplished by Node  $i$  and subsequently, accomplished by Node  $j$ , as well.

The computation of influence weights, is followed by averaging them to distinguish those nodes that are deemed highly active, based on historical logs data. This process reveals the Influence Weights fuzzy set membership function, which is formulated as follows:

$$InfluenceWeightsAvg_i = \frac{\sum_{j=1}^n InfluenceWeights_{ij}}{n}. \quad (4)$$

Equation (4) shows that the cumulated value of influence weights that a specific Node  $i$  has over each Node  $j$  across the network, is normalized by the size of the social network  $n$ .

### 4.2.3 Fuzzy decision making

After computing both central and influential user fuzzy sets for every user, we decide on key users using a fuzzy-set intersection operation for every Node  $i$ , considering the corresponding fuzzy sets  $CentralityWeight_i$  and  $InfluenceWeightsAvg_i$ , using the following formulation:

$$Intersection_i = \min(CentralityWeight_i, InfluenceWeightsAvg_i). \quad (5)$$

Equation (5) shows that the *intersection* between the fuzzy sets picks the smallest of degree centrality and influence weight values. Subsequently, Eq. (6) shows that, ultimately the key users are those which maximise their intersecting structural and influence fuzzy-membership sets. The rationale of this approach is to address deficiencies each node may have in either its structural or influence power dimensions. This is why, the fuzzy-intersection considers the minimum of both values, so that users with less deficiency in either attribute get picked. This results in a set of user nodes with a single associated value, that is the least deficient, in terms of

influence or structural shortcomings. Subsequently, the key nodes are determined based on the maximum of these single valuations of each node, as formulated by Eq. (6):

$$KeyUsers = \operatorname{argmax}_{i \in N}(Intersection_i). \quad (6)$$

In Eq. (6),  $N$  represents the social network user nodes set.

### 4.2.4 Seed set

At last, the seed set which represents the actual set of users to drive influence propagations is selected as the top  $k$  nodes of key users, where  $k$  is a parameter that depends on the allocated budget to a given marketing campaign to account for cost involving in recruiting seed set users to promote a given product or a service or spread a desirable behavioural campaign, such as stop-smoking. Hence, we need to rank the key users in order to be able to pick the top  $k$  ones. For that, Algorithm 3 is employed to evaluate the influence spread for each user using the IC model. For each run of the algorithm, we account the number of activations that a candidate key user scores. The computational cost of this approach is similar to that of IC model, however the input key user nodes are judiciously picked in our case using our proposed fuzzy-logic based selection process. Our approach also contrasts with LTM which does not consider action logs data, like we advocate. In addition, LTM is NP-Hard, calling for heuristic approaches to harness the problem. Instead, we harness the problem through the gradual three modular steps process that are: (1) detecting virtual communities using correlations between user actions, (2) identifying key users in each of these communities, and (3) finding the seed-set (among those key users) to propagate influence across the entire social network.

---

#### Algorithm 3 Seed-Set of Highly Influential Nodes

---

1. Set  $InfluenceThreshold \leftarrow$  Nodes activation probability
  2. For  $\forall i \in KeyUsers$  do
    - (a) At timestamp  $t = 0$ , activate  $i \in KeyUsers$  and add it to  $Coverage_0$
    - (b) At subsequent timestamps  $t > 0$ , For  $\forall i \in Coverage_{t-1}$  do
      - i. For  $\forall j_{inactive}$  if  $InfluenceWeight_{i,j} \geq InfluenceThreshold$ 
        - A. Activate  $j$
        - B.  $ActiveList = ActiveList \cup \{j\}$
        - C.  $TotalCoverage = TotalCoverage + 1$
      - ii. All activated nodes at this step are included in  $Coverage_t$
      - iii. At timestamp  $t$ , the process stops if  $Coverage_t = 0$  /\*no further nodes can be activated\*/
  3. Augment  $S$  with nodes  $i$  having highest  $TotalCoverage$
  4. Return  $S$
-

## 5 Experiments and performance analysis

This section describes the experiments we conducted to evaluate the community-based influence propagation approach we introduced in this paper. As mentioned in Step 1 of Algorithm 2, we propose to use a similarity based preprocessing step to enrich the input social network using Algorithm 1, before applying Step 2 which detects communities in the enhanced social network. The resulting Similarity-CNM task is poised to detect better community structure as explained further in Sect. 4.1. Hence, we propose to first reveal the outcomes of this pre-processing step, whereby community quality is measured using Modularity as evaluation metric. Subsequently, we implemented the remaining steps of Algorithm 1 to generate the key-users which accumulate both a favorable location in the network and a good account of influence ( $S$ ). And finally, we run the second experiment to assess the propagation extent of influence propagated by the highest key-users, which form the actual seed-sets. Throughout both experiments, we hypothesize that the similarity based preprocessing step on a social network  $G$  is effective with respect to the quality of communities, which are used to select candidates for the further influence-propagation step. In doing so, we hypothesize also that the fuzzy-logic based combination of favourable location within those communities, and the prior activity history elects highly influential candidates across the entire network. We implemented the proposed algorithms in this paper using Python and related iGraph and Networkx libraries. A Mac OS X version 10.14.2 (Mojave) platform powered with an i7 processor of 2.50GHz and a RAM of 16 GB was used to implement the algorithms presented in this paper in order to evaluate their performance.

### 5.1 Datasets

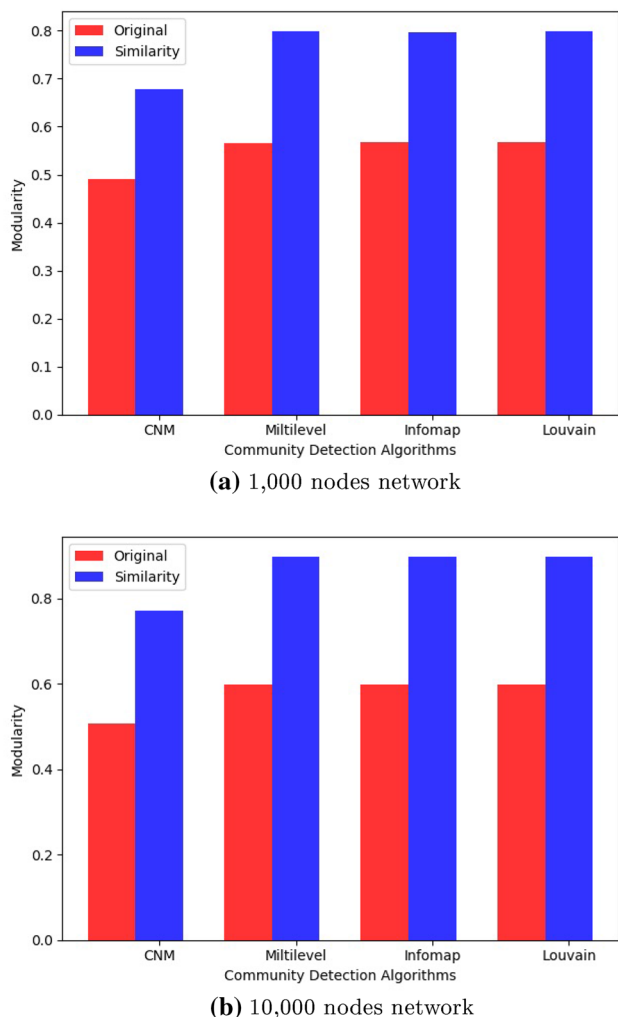
We used both artificially-generated social networks and actual network data. To evaluate the similarity-based community detection algorithm, we used LFR benchmark networks as dataset (Lancichinetti 2008). This benchmark was used in several researches dealing with community-detection in social networks (Cao et al. 2015; Hafez et al. 2014; Chen et al. 2016; Emmons et al. 2016; Orman et al. 2012). Simulated networks are employed in the community-detection experiment to overcome the difficulty to evaluate communities in real-world networks due to an absence of community ground-truths (Cao et al. 2015), and to assess community-quality under varying degrees of structural parameters. Nevertheless, LFR Benchmark networks do simulate networks that are very close to real-world social networks' data (Bródka et al. 2010), and this benchmark is becoming a de-facto standard network-generator for evaluating the

performance of different community-detection algorithms (Largerone et al. 2015). We generated a network of 10,000 nodes using LFR benchmark for the first experiment. The most important parameter used to vary the structure of the network is known as the mixing parameter  $\mu$ , which represents the fraction of intra-community edges incident to each node. Its value ranges from 0 to 1, where 0 results in graphs that have high community structure, and 1 results in graphs that have low community structure. The mixing parameter generates this connection based on  $(1 - \mu)$  for intra-community edges and  $(\mu)$  for inter-community edges. Thus, values between 0 and 0.5 yield proper community structures, and values between 0.5 and 1 yield loose community structures. The other parameters are  $\tau_1$  and  $\tau_2$ , respectively the "power law exponent of degree distribution" and "power law exponent for the community size distribution" (Lancichinetti 2008; Lancichinetti and Fortunato 2011), and which are respectively set to 2 and 1.5 in our experiment. Further parameters are the average and maximum node degree set to 10 and 50 respectively in our experiment, and the community-size set between 20 and 60. These values are consistent with those proposed by LFR benchmark providers (Lancichinetti 2008; Lancichinetti and Fortunato 2011).

For the second batch of experiments however, related to influence-propagation reach, we employed real-world data sets from Flickr social network. This social-network is distinguished by photo sharing activities. Users of Flickr post photographs or include them into blogs and other users may "like" the posted photographs as an instance of an activation. This dataset is graciously made available by some published works (Cha et al. 2009), and consists of over 2.5 million nodes with over 33 million links. Due to computational constraints and as part of our preliminary experiments, we extracted two subgraph samples of 500 and 5000 nodes, randomly to observe the results of our experiments across real networks and assess the scalability properties of the obtained results. The targeted indicators from this second batch of experiments relates to the performance of the fuzzy-logic based intersection between favourable and influential nodes, that are poised to optimize diffusion across social networks. The extracted subgraphs preserve the original links which amount to 26,223 edges for the 500 nodes network and 242,600 edges for the 5000 nodes network.

### 5.2 Candidate algorithms

In the context of community detection, we propose to illustrate the performance of the similarity-based algorithm against a series of known community-detection algorithms, including pioneering CNM (Newman 2004), as well as InfoMap (Rosvall and Bergstrom 2008), Louvain (Blondel et al. 2008), and Multilevel (Rotta and Noack 2011) algorithms. CNM is modularity-based and very fast. Infomap is a search



**Fig. 4** Modularity of community structure when applying similarity network

algorithm for minimizing a map equation over possible network partitions. Louvain is a greedy optimization approach that maximizes the modularity of a partition in the network in two steps. Initially, “small” communities are established through a local optimisation of the modularity value, and then community nodes are aggregated to construct a new network. The process iterates over these two steps until a maximum value of the modularity is reached. The multilevel refinement method is a multistep approach, which repeatedly prioritises the process of joining pairs of clusters that do not decrease the modularity. The priority criterion is a parameter of the algorithm.

Subsequently, and in the context of influence diffusion, we conducted experiments to measure the performance gain of our proposed approach compared to the original IC approach (Kempe et al. 2003), by setting the input set of triggering nodes  $A_0$ . The members of this set should be carefully selected to maximise influence propagation. IC model does

not exploit action correlations among users of the social network whereas our proposed Algorithm 3 integrates influence weights based on common actions weights. In addition, IC propagates influence over the input network, whereas we consider our enriched similarity-network. To assess the gain in activated nodes following influence propagation, we apply IC model with randomly selected seed-nodes against the seed-set users generated from 3. With this comparison, we evaluate the value of the fuzzy-intersection introduced in this paper to select the most appropriate nodes for influence diffusion.

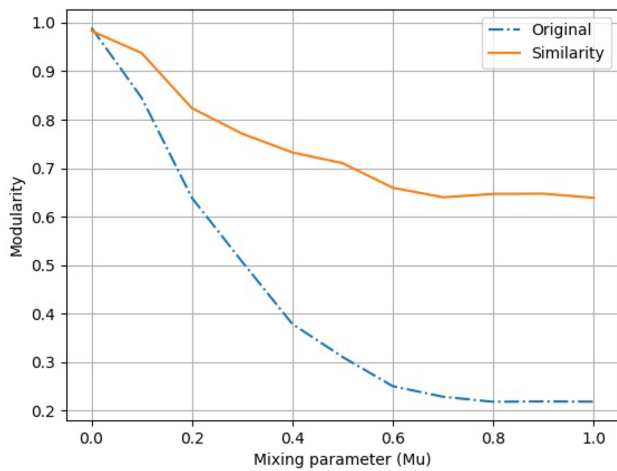
### 5.3 Performance metrics

A practical metric used to evaluate the fabric structure of communities is the modularity, which is denoted  $Q$  and which evaluate the partitions in a social network. This evaluation is based on the variance of the amount of edges linking nodes within the same cluster from an expected amount of edges in an arbitrary network (that is typically unstructured). Better communities are detected when this difference is large. According to (Clauset et al. 2004a), the value of  $Q$  above 0.3 is considered as a significant community structure. This value is derived from the following formula:

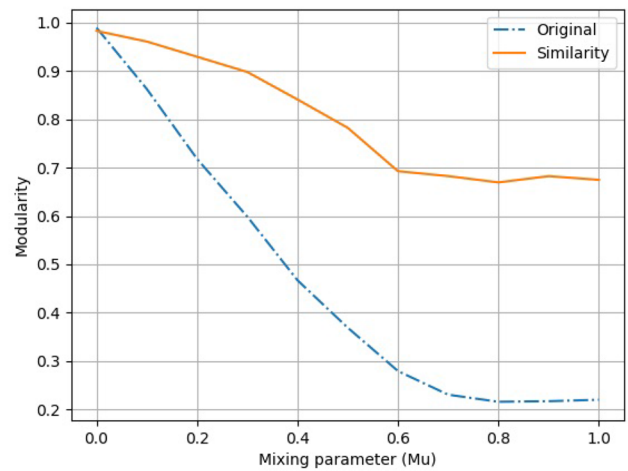
$$Q = \sum_i (e_{ij} - a_i^2), \quad (7)$$

where  $e_{ij}$  is the proportion of edges linking vertices in communities  $i$  and  $j$ , and  $a_i = \sum_j e_{ij}$  is the proportion of edge-endpoints that connect vertices in community  $i$ .  $Q$  value ranges between  $[-1, 1]$ , and measures the density of vertices within the same community to that of nodes belonging to a different community. The larger the modularity score, the better is the partitioning of nodes into communities. A low-score means there is less community structure and high-score means communities are very well partitioned (structured). The similarity threshold of Algorithm 1 was set to 0.005 following observed experimental instances of  $Q$  for various threshold settings.

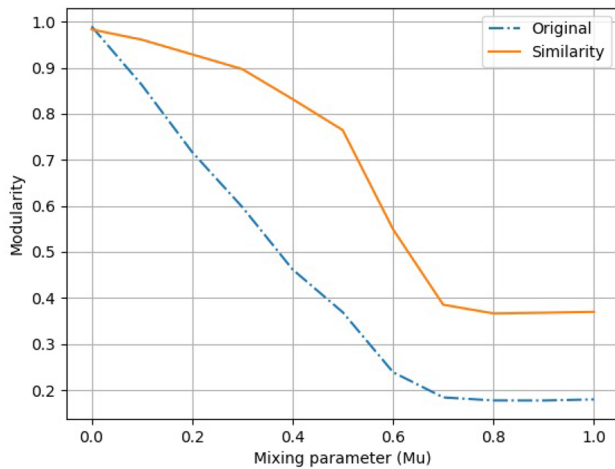
As for the influence-maximization evaluation, we made use of activated nodes size as a performance metric. We determine the number of activated nodes as it is the paramount element to contrast the performance of influence propagation for both the IC model and our proposed fuzzy-logic based model. Nodes activation can be explained as the embracement of a certain action by a node, which is triggered by another node. This is the main goal for influence propagation, whereby algorithms strive to scale-up nodes activation that conveys the adoption of a marketed product or a targeted behaviour. The influence threshold of Algorithm 3 was set to 0.1.



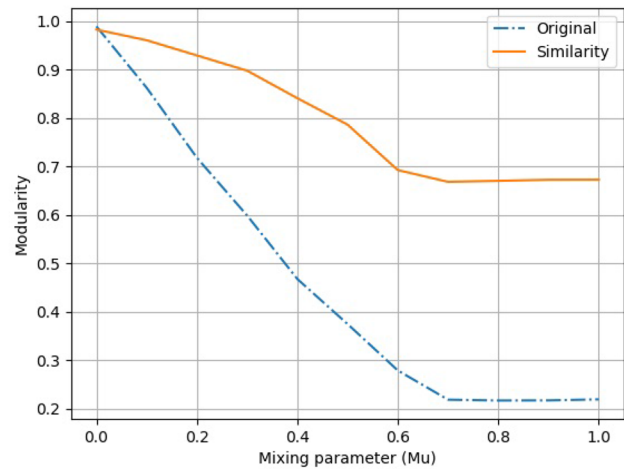
(a) Original vs. Similarity CNM



(b) Original vs. Similarity Multilevel



(c) Original vs. Similarity Infomap



(d) Original vs. Similarity Louvain

Fig. 5 Community-detection modularities over 10,000 node networks with various topologies

### 5.4 Performance results

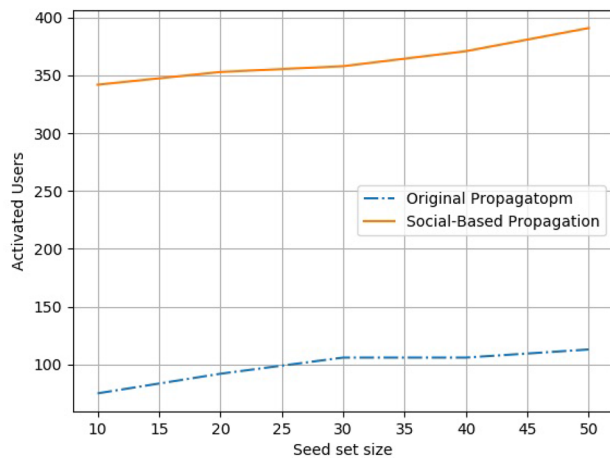
Our experiment results are organized in two stages. First, we show the gain in modularity of communities obtained when applying our pre-processing step of Algorithm 1. Then, we reveal the gain in activated nodes obtained when our proposed fuzzy-logic based approach to elicit key-users shown in Algorithm 2 is employed, and subsequent diffusion of Algorithm 3 is carried out to generate the seed set of highly-influential users.

#### 5.4.1 Similarity-based preprocessing

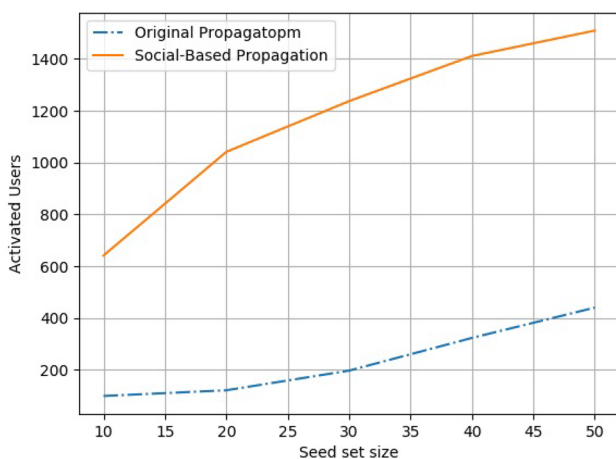
Figure 4a shows the performance when our proposed enriched synthetic-network is used for community-detection by a range of landmark algorithms. The original

LFR-generated network of 1000 nodes delivers a lower community structure than the enriched similarity network obtained by adding edges between similar nodes as dictated by Algorithm 1. The same 20% margin-gain in modularity is observed in a higher-scale network of 10,000 nodes as shown in Fig. 4b. Another observation is that the similarity network creates better opportunity for community clusters, when the network size increases as the modularity rises between the 1000-nodes network to the 10,000-nodes network, whereas it appears stationary in the original network cases. In both experiments, the mixing parameter  $\mu$  was set to 0.3.

From the above experiment, we observe the value of the judicious similarity-function based on common-neighbors employed to support community-detection algorithms detect better communities. We use the synthetic similarity network



(a) 500 Flickr network nodes after 10 diffusion steps



(b) 5,000 Flickr network nodes after one diffusion step

Fig. 6 Activated Flickr network nodes with varying seed-set sizes

to detect virtual-communities via various community detection algorithms, for the purpose of using the enhanced community-structure to find the important nodes within each of these virtual communities. Later, in subsequent experiments, we observe the value of the proposed fuzzy-logic based approach to calculate the influence weight for each node within each community, based on both centrality measure and common actions history. Using the community structure scales-down the complexity of finding important nodes.

The enhanced community-structure brought about by the similarity-network spans various network topologies, as illustrated further in Fig. 5, where similarity-based community detection algorithms outperform original algorithms across a range of mixing parameter values. The performance of detected community-structures degrades as the value of mixing parameter  $\mu$  rises. Each point in the graphs represents an instance of modularity for a given LFR-generated network of 10,000 nodes shaped through the indicated mixing parameter in the x-axis. A low mixing-parameter

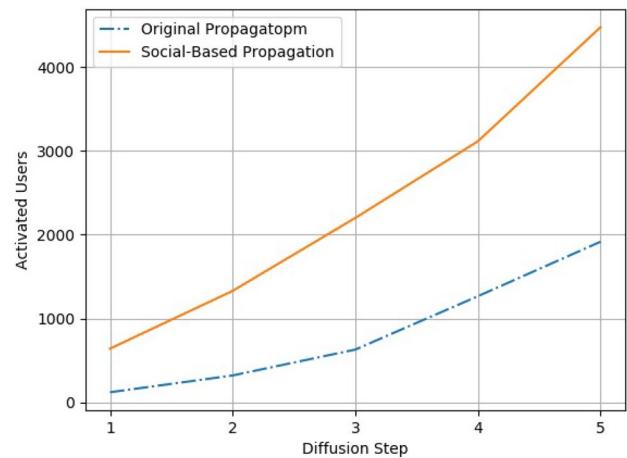


Fig. 7 Activated nodes with varying diffusion steps using 10 seed nodes in a 5000-nodes Flickr network

is conducive to dense community-structures, since the fraction of neighboring nodes outside any community (i.e.  $\mu$ ) is low, and hence higher modularity is inferred given the tight community structures of the sample network. On the other hand, a high mixing-parameter is conducive to loose community-structures since the fraction of nodes outside any community is high, and hence a low modularity is inferred. However, for each community-detection algorithm, the degradation is moderate when applying our proposed preprocessing approach to the original network.

Furthermore, for each community-detection algorithm, the modularity does not fall below the 0.3 threshold, which is the minimum value mentioned in Clauset et al. (2004a), to indicate a significant community structure. It is noteworthy that InfoMap case shown in Fig. 5c, the similarity-network shows the steepest loss in modularity. This is because Infomap community-detection algorithm is distinguished from the other community-detection algorithms as it relies on a map equation, whereas the others are modularity-maximisation approaches. The employed map equation in Infomap partitions the network following some patterns within the network, to build communities. Hence, for our subsequent step of influence maximization, we use a candidate from modularity-based approaches, namely CNM.

#### 5.4.2 Social-influence propagation

Through our second batch of experiments, we evaluate the influence-propagation reach when employing the fuzzy-logic based approach discussed in Sect. 4.2, after preprocessing the social network using the approach presented in Sect. 4.1. As stated earlier, real-world data sets from Flickr social network are used in these experiments. Two subgraph samples of 500 nodes with 26,223 edges, and 5000 nodes

with 242,600 edges, have been extracted from Flickr data to evaluate the diffusion spread and the scalability performance of candidate algorithms. The output is measured in terms of the number of activated nodes. This performance metric estimates the diffusion along the network, given an input of judiciously selected seed-set users, which in our proposed approach are inferred from our fuzzy-logic based technique.

The results produced from these experiments batch reveal interesting tradeoffs between our proposed method and IC model, whereby a larger number of nodes are activated by our proposed community-based influence-diffusion model (discussed in Algorithm 2). and yet involving a smaller seed-set (compared to IC model). This is illustrated by Fig. 6 which reports the influence diffusion results for two sample-size Flickr networks. First, Fig. 6a shows the results for a snapshot of 500 nodes, where our proposed social-influence based propagation which uses the fuzzy-logic discrimination to identify seed-set nodes quickly reaches a high range of nodes activation while using few seed nodes. Indeed, 10 nodes activate about 350 nodes in our social-based propagation, while original IC model which chooses random seed-set nodes activates about 30 nodes, only. As the seed-set threshold increases. the social-influence approach scales up the range of activated nodes to the higher-end, reaching about about 400 nodes for an initial 50 seed nodes. The results show the pursuit of nodes activation to cover almost the entire 500 nodes of the sample Flickr social network. They also show that IC model requires 50 seed nodes to activate the third of what our proposed social-influence propagation achieve with less merely 10 nodes. The activation gain is about 90% attributed to our proposed social-influence approach compared to IC benchmark. These results are the outcome of 10 diffusion steps, where the judicious combination of seed nodes' location and historical influence brought by the fuzzy intersection of Eq. (5) enable faster influence propagation to the high-end of the social network. The results show that exploiting correlations exposed by centrality attribute and historical action logs, the influence propagation process scales higher the activation process. The triggering users are more successful in persuading neighbours or neighbours-of-neighbours to embrace the propagated action.

In another experiment, the size of the Flickr sample network is extended to 5000 nodes. The results are illustrated in Fig. 6b and in this case we report the results of just one diffusion step, and limiting the threshold of "important users" selected for their centrality value within their communities to 10. This is the *centralityThreshold* mentioned in Sect. 4.2.1, which reflects the topological eligibility of seed-set nodes. Figure 6b shows that in just one diffusion step, the number of activated nodes rises quickly to over 600 users, in an influence campaign driven by just 10 seed users. By contrast, IC model activates about 100 users with the same number of seed nodes. However, the gap

between the two models grows when increasing the seed-set size to reach 1000 more activated nodes by the social-based propagation over original IC model. Hence, using IC approach, 400 nodes are activated by the 50 nodes of the seed set used to diffuse influence in the social network as shown in Fig. 6c, whereas this number climbs to over 1400 in the social propagation model. This is an important outcome, considering investment decisions made by businesses to promote a product using our approach, as they could persuade less number of initial people to promote their product to expand the outcome of a marketing campaign. This approach induces substantial marketing savings to provide free samples to those influence-inceptive individuals forming the seed-set. In addition, businesses raise their income, as those inceptive-individuals have the capacity to entice a large number of social-network users to adopt the product at a later stage.

We also noticed some more interesting results when analyzing the performance with varying diffusion steps as illustrated in Fig. 7. The social-based propagation is able to reach almost the entire Flickr social network of 5000 nodes after just five diffusion steps using only 10 seed-set nodes, and outperforming the original IC propagation by an increasing margin as the diffusion step rises. This means that 0.2% of individuals in a social network are able to convince the entire network population. This influence scalability is attributed to the fuzzy-set intersection approach we employed to determine key users. The combined value obtained from both *CentralityWeight<sub>n</sub>* and *InfluenceWeightsAvg<sub>n</sub>* criteria to decide on key user node *n*, compensates the defects in either centrality or influence weight to reach neighbouring nodes. The impacts of this result contribute to an efficient seed set of potential candidates for influence propagation, by its reduced size and higher influence (reachability).

## 6 Conclusion

In the presented work within this paper, we address the prominent social-network problem pertaining to influence-maximization, for which we contribute a computational-intelligence approach to expand the influence diffusion rates in contemporary social networks. The experimental analysis results reveal the potential benefits of using a community-enrichment preprocessing step before applying influence-diffusion algorithms. We also suggested a new method to find "key nodes" in social networks using a computational intelligence approach that adapts fuzzy-logic theory to key users selection. This technique discovers the most influential nodes as seed set for influence propagation, by combining multiple criteria such as nodes' location and influence weights in the social network. The propagation of influence

in social networks involves naturally some vagueness, given dynamic nodes' relationships and location in the network. The proposed fuzzy-logic approach is suggested to overcome this typical vagueness in social networks, by combining both of these node properties to assert key nodes that are candidate seed set members for diffusing influence. These correlations have practical implications across a range of business, political or social campaigns that aim at generating revenues while minimizing costs, or adopting desired behaviours across a society with fewer interventions.

Future directions to extend the influence maximization algorithm presented in this paper are numerous to investigate further efficiency and scalability opportunities. We are also working on applying the proposed approach to other real-word datasets such as YouTube and come up with new insights about robustness in finding the most influential seed set. We are also exploring the effectiveness of employing various centrality attributes for a better precision of the obtained results, such as betweenness, closeness, etc.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Acar A, Muraki Y (2011) Twitter for crisis communication: lessons learned from japan's tsunami disaster. *Int J Web Based Communities* 7(3):392–402
- AlFalahi K, Atif Y, Harous S (2013) Community detection in social networks through similarity virtual networks. In: Proceedings of the 4th business applications of social network analysis workshop, (BASNA 2013) in conjunction with the IEEE and ACM, (ASONAM 2013), ACM
- AlFalahi K, Atif Y, Abraham A (2014) Models of influence in online social networks. *Int J Intell Syst* 29(2):161–183
- Arenas A, Fernández A, Gómez S (2008) Analysis of the structure of complex networks at different resolution levels. *N J Phys* 10(053):039
- Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. In: Proc 12th ACM SIGKDD international conference on knowledge discovery and data mining
- Baig F, Ashraf MW, Ahmed Z, Imran M, Tayyaba S, Khan MS (2013) Design and simulation of fuzzy logic based elid grinding control system. *Int J Adv Technol Eng Res (IJATER)* 3(1):79–88
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 10:10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>. arXiv: 0803.0476
- Bródka P, Musiał K, Kazienko P (2010) A method for group extraction in complex social networks. Knowledge management, information systems, e-learning, and sustainability research. Springer, Berlin, pp 238–247
- Cao X, Wang X, Jin D, Guo X, Tang X (2015) A stochastic model for detecting overlapping and hierarchical community structure. *PLoS One* 10(3):e0119171
- Cha M, Mislove A, Gummadi KP (2009) A measurement-driven analysis of information propagation in the flickr social network. In: Proceedings of the 18th international world wide web conference (WWW09)
- Chen Y, Zhao P, Li P, Zhang K, Zhang J (2016) Finding communities by their centers. *Sci Rep* 6:24017. <https://doi.org/10.1038/srep24017>
- Clauset A, Newman MEJ, Moore C (2004a) Finding community structure in very large networks. *Phys Rev E* 70(6):066111
- Clauset A, Newman MEJ, Moore C (2004b) Finding community structure in very large networks. *Phys Rev E* 70(066):111. <https://doi.org/10.1103/PhysRevE.70.066111>
- Cosley D, Huttenlocher DP, Kleinberg JM, Lan X (2010) Sequential influence models in social networks. In: ICWSM
- Domingos P, Richardson M (2001) Mining the network value of customers. In: Proc 7th ACM SIGKDD international conference on knowledge discovery and data mining, pp 57–66
- Dumenco S (2011) A very brief (cartoon) history of social influence. <http://adage.com/article/digitalnext/a-cartoon-history-social-influence-2006-2011/226942/>. Accessed 6 June 2018
- Emmons S, Kobourov S, Gallant M, Börner K (2016) Analysis of network clustering algorithms and cluster quality metrics at scale. *PLoS One* 11(7):e0159161
- Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3):1–75
- Fortunato S, Latora V, Marchiori M (2004) Method to find community structures based on information centrality. *Phys Rev E* 70(5):056104
- Goyal A, Bonchi F, Lakshmanan LVS (2010) Learning influence probabilities in social networks. In: Proceedings of the third ACM international conference on web search and data mining, WSDM '10, pp 241–250
- Hafez AI, Hassanien AE, Fahmy AA (2014) Testing community detection algorithms: a closer look at datasets. Social networking. Springer, Cham, pp 85–99
- Hanneman RA, Riddle M (2005) Introduction to social network methods. Centrality and power, Chap (10). Department of Sociology at the University of California, Riverside
- Hellmann M (2001) Fuzzy logic introduction. <http://epsilon.nought.de/tutorials/fuzzy/fuzzy.pdf>. Accessed 6 June 2018
- Hosseini-Pozveh M, Zamanifar K, Naghsh-Nilchi AR (2017) A community-based approach to identify the most influential nodes in social networks. *J Inf Sci* 43(2):204–220. <https://doi.org/10.1177/0165551515621005>
- Kahraman C (2008) Multi-criteria decision making methods and fuzzy sets. In: Kahraman C (ed) Fuzzy multi-criteria decision making, Springer optimization and its applications, vol 16. Springer, New York, pp 1–18
- Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: Proc 9th ACM SIGKDD international conference on knowledge discovery and data mining, pp 137–146
- Khousa EA, Atif Y (2018) Social network analysis to influence career development. *J Ambient Intell Humaniz Comput* 9(3):601–616. <https://doi.org/10.1007/s12652-017-0457-9>
- Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. *Science* 311(5757):88–90
- Kundu S, Pal SK (2015) FGSN: fuzzy granular social networks: model and applications. *Inf Sci* 31:100–117
- Lancichinetti A (2008) Benchmark graphs for testing community detection algorithms. *Phys Rev E* 78(4):89. <https://doi.org/10.1103/PhysRevE.78.046110>



- Lancichinetti A, Fortunato S (2011) Limits of modularity maximization in community detection. *Phys Rev E* 84(6):066,122. <https://doi.org/10.1103/PhysRevE.84.066122>
- Largerion C, Mouguel PN, Rabbany R, Zaïane OR (2015) Generating attributed networks with communities. *PLoS One* 10(4):54–62
- Li Z, Zhang S, Wang RS, Zhang XS, Chen L (2008) Quantitative function for community detection. *Phys Rev E* 77(3):36–109
- Nair PS, Sarasamma ST (2007) Data mining through fuzzy social network analysis. In: NAFIPS 2007–2007 annual meeting of the North American Fuzzy Information Processing Society, pp 251–255. <https://doi.org/10.1109/NAFIPS.2007.383846>
- Newman ME (2004) Fast algorithm for detecting community structure in networks. *Phys Rev E* 69(6):066133. <https://doi.org/10.1103/PhysRevE.69.066133>
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69(2):026113
- Newth DJ (2006) The structure of social networks. In: Perez P, Batten DF (eds) *Complex science for a complex world: exploring human ecosystems with agents*. ANU E Press, Canberra, pp 71–94
- Orman GK, Labatut V, Cherifi H (2012) Qualitative comparison of community detection algorithms. CoRR [arXiv:1207.3603](https://arxiv.org/abs/1207.3603)
- Peneva V, Ivan P (2008) Multicriteria decision making based on fuzzy relations. *Cybern Inf Technol* 8(4):3–12
- Rahman SM, Ratrouf NT (2009) Review of the fuzzy logic based approach in traffic signal control: prospects in Saudi Arabia. *J Transp Syst Eng Inf Technol* 9(5):58–70
- Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. In: *Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '02)*, pp 61–70
- Rojas R (1996) *Neural networks systematic introduction*. Springer, New York
- Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci* 105(4):1118–1123. <https://doi.org/10.1073/pnas.0706851105>. <http://www.pnas.org/content/105/4/1118.full.pdf>
- Rotta R, Noack A (2011) Multilevel local search algorithms for modularity clustering. *J Exp Algorithmics* 16:2–3. <https://doi.org/10.1145/1963190.1970376>
- Scott J (2000) *Social network analysis: a handbook*. SAGE, Beverly Hills
- Shi X, Zhu J, Cai R, Zhang L (2009) User grouping behavior in online forums. In: *Proc 15th ACM SIGKDD international conference on knowledge discovery and data mining*, pp 777–786
- Solis B (2011) The interest graph on twitter is alive: studying Starbucks top followers. <http://bitly.com>. Accessed 6 June 2018
- Sun J, Tang J (2011) *Social network data analysis*. Springer Science+Business Media, New York
- Tang J, Sun J, Wang C, Yang Z (2009) Social influence analysis in large-scale networks. In: *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '09)*, pp 807–816
- Wang Y, Cong G, Song G, Xie K (2010) Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In: *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, ACM, New York, NY, USA, KDD '10*, pp 1039–1048. <https://doi.org/10.1145/1835804.1835935>
- Wolfram (2014) Fuzzy logic, example 7: choosing a job. <http://www.wolfram.com/products/applications/fuzzy.logic/examples/job.html>. Accessed 6 June 2018
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353
- Zadeh LA (1984) Making computers think like people. *IEEE Spectr* 8:26–32

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.