

RESEARCH

Open Access



Exploring the step function distribution of the threshold fraction of adopted neighbors versus minimum fraction of nodes as initial adopters to assess the cascade blocking intra-cluster density of complex real-world networks

Natarajan Meghanathan*

*Correspondence:
natarajan.
meghanathan@jsums.edu
Jackson State University,
Jackson, MS 39217, USA

Abstract

We first propose a binary search algorithm to determine the minimum fraction of nodes in a network to be used as initial adopters (f_{IA}^{\min}) for a particular threshold fraction (q) of adopted neighbors (related to the cascade capacity of the network) leading to a complete information cascade. We observe the q versus f_{IA}^{\min} distribution for several complex real-world networks to exhibit a step function pattern wherein there is an abrupt increase in f_{IA}^{\min} beyond a certain value of q (q_{step}); the f_{IA}^{\min} values at q_{step} and the next measurable value of q are represented as $\underline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$ respectively. The difference $\overline{f_{IA}^{\min}} - \underline{f_{IA}^{\min}}$ is observed to be significantly high (a median of 0.44 for a suite of 40 real-world networks studied in this paper) such that we claim the $1 - q_{step}$ value (we propose to refer $1 - q_{step}$ as the Cascade Blocking Index, CBI) for a network could be perceived as a measure of the intra-cluster density of the blocking cluster of the network that cannot be penetrated without including an appreciable number of nodes from the cluster to the set of initial adopters (justifying a relatively larger $\overline{f_{IA}^{\min}}$ value).

Keywords: Information cascade, Cascade capacity, Blocking cluster, Intra-cluster density, Initial adopters, Step function distribution, Binary search

Introduction

Information cascade in complex networks is a phenomenon by which one or more nodes in a network adopt a decision based on the decision made by their neighbor nodes. Information cascade in complex networks is typically initiated through a set of nodes called the *Initial Adopters* (considered to have adopted a particular decision) and the cascade proceeds in a sequence of iterations. We follow a threshold-based complex contagion model (Watts 2002) for sequential information cascade: A node (that has hitherto not adopted a decision) adopts a decision in a particular iteration if the fraction of its

neighbors who have adopted the decision is greater than or equal to a threshold fraction (q). The iterations stop if all the nodes have adopted some decision or no new node adopts a decision in the latest iteration. An information cascade is said to be "complete" (Easley and Kleinberg 2010) if all the nodes in the network arrive at a unanimous decision starting with a set of initial adopters who also adopt the same decision. We focus on *complete information cascade* (also referred to as global cascade in the literature) in this paper. For a given set of initial adopters, the largest possible value for the threshold fraction (q) of adopted neighbors that would lead to a complete information cascade is called the *cascade capacity* of the network (Easley and Kleinberg 2010). The larger the size of the set of initial adopters and more central are the initial adopter nodes, the larger the cascade capacity of a network. We use nodes with larger degree centrality (DEG) (Newman 2010) or betweenness centrality (BWC) (Freeman 1977) as the nodes that are part of the set of initial adopters. These two centrality metrics have been observed (Jalili and Perc 2017) to be effective in speeding up the adoption process for the rest of the nodes in the network.

Clusters are perceived as the major bottleneck for complete information cascade (Easley and Kleinberg 2010). A cluster in a network is a subset of the nodes that have more links among themselves compared to links with the rest of the nodes in the network (Newman 2010). The intra-cluster density of a cluster is a measure of the number of links connecting two nodes within the cluster and the inter-cluster density of a cluster is a measure of the number of links connecting the nodes in the cluster to nodes outside the cluster. The intra-cluster density of a cluster is computed as the minimum of the intra-cluster densities of the bridge nodes (nodes that have edges with nodes both inside and outside the cluster) of the cluster (Easley and Kleinberg 2010). The intra-cluster density of a bridge node is the ratio of the number of neighbors inside the cluster and the total number of neighbors.

We begin our research in this paper to find a solution for the following problem: For a given threshold fraction (q) of the adopted neighbors that would result in a complete information cascade, what is the minimum number of nodes in a network that need to be used as the initial adopters (IA_{min})? An efficient binary search approach for the above problem is not currently available or used in any of the related works in the literature. With a search space of $(0, \dots, N]$, where N is the number of nodes in the network, we propose a binary search algorithm to find the actual value of IA_{min} for a network for a given q . The minimum fraction of nodes as initial adopters (f_{IA}^{min}) corresponding to IA_{min} is then computed as IA_{min}/N . Using this binary search algorithm, we proceed further and build a distribution of the q versus f_{IA}^{min} values. To our surprise, we observe the q versus f_{IA}^{min} distribution to exhibit a step function pattern for 37 of the 40 complex real-world networks analyzed in this research, wherein there is an abrupt increase in f_{IA}^{min} beyond a certain value of q (q_{step}); the f_{IA}^{min} values at q_{step} and the next measurable value of q are represented as $\underline{f_{IA}^{min}}$ and $\overline{f_{IA}^{min}}$ respectively. We developed a second binary search algorithm (that makes use of the algorithm to determine IA_{min}) to determine the actual value of q_{step} for each of the 40 complex networks and the centrality metrics: DEG and BWC. The differences between the $\underline{f_{IA}^{min}}$ and $\overline{f_{IA}^{min}}$ values were observed to be significantly high for several complex networks. We claim that the value of " $1 - q_{step}$ " for a network could be perceived as a measure of the intra-cluster density of the blocking cluster

of the network that could not be penetrated unless we include one or more nodes of this blocking cluster as part of the set of initial adopters (justified by the sharp increase in the f_{IA}^{\min} values from $\underline{f_{IA}^{\min}}$ to $\overline{f_{IA}^{\min}}$ in the vicinity of q_{step}).

A cluster is referred to as a blocking cluster if the fraction of adopted neighbors for the bridge nodes of the cluster is less than the threshold fraction (q) of adopted neighbors and information cascade cannot penetrate through such a cluster. The sharp increase in the f_{IA}^{\min} values from $\underline{f_{IA}^{\min}}$ to $\overline{f_{IA}^{\min}}$ in the vicinity of q_{step} for a network suggests that there should be at least one such blocking cluster (of intra-cluster density measured in the form of $1 - q_{step}$) for which starting with $\underline{f_{IA}^{\min}}$ fraction of initial adopters is not sufficient and one or more nodes (typically the bridge nodes and some internal nodes) from this cluster need to be part of the set of initial adopters (corresponding to the $\overline{f_{IA}^{\min}}$ fraction) to accomplish complete information cascade. We propose that " $1 - q_{step}$ " be called the Cascade Blocking Index (CBI) of a network, a quantitative measure of the difficulty in penetrating through the blocking cluster(s) of the network. The larger the CBI value for a network, the larger the intra-cluster density of the blocking clusters of the network and vice-versa. While a lower CBI value is preferred for positive information to seamlessly get adopted by the nodes in the network, a larger CBI value is preferred for a network to keep away the epidemics/pandemics from infecting and spreading through the nodes.

As information cascade is observed to be more successful in clustered networks rather than non-clustered networks (Ikeda et al. 2010), the problem of inferring community structures of complex networks solely on the basis of information cascade (i.e., the underlying network structure is not known) gained attention in the literature (Rodriguez et al. 2014; Ramezani et al. 2018; Prokhorenkova et al. 2019) in recent times. To the best of our knowledge, there has been no work in the literature to quantify the maximum (or the minimum or the distribution) of the intra-cluster densities of the clusters without actually determining the clusters. The pioneering book chapter (# 19) in Easley and Kleinberg (2010) relates intra-cluster density with the threshold fraction (q) of adopted neighbors and refers to a cluster as a blocking cluster if its intra-cluster density exceeds $1 - q$. Other than this, there is no other work reported in the literature that relates intra-cluster density with the threshold fraction of adopted neighbors for information cascade. The CBI value for a network (proposed in this paper) could be considered as an upper bound for the intra-cluster density (also a measure of the clusterability of the network) that could be expected of the clusters determined by the community detection algorithms.

The proposed methodology definitely has advantages over clustering, because with the latter approach: one has to first determine all the clusters of a network and then identify the blocking cluster (the cluster with the largest intra-cluster density) of the network. There are a multitude of clustering algorithms in the literature, each with different time complexities as well as the output (the set of clusters) of one clustering algorithm might be different from another for the same graph. In Sect. 5.2, we demonstrate a strong correlation (both rank-wise and prediction-wise) between the CBI values and the largest of the intra-cluster density values (with the clusters determined using the Louvain algorithm) of the real-world networks. We thus claim our work of quantifying the intra-cluster density of the blocking cluster for information cascade (without running any clustering algorithm) by using the threshold fraction of adopted neighbors and the

minimum fraction of nodes as initial adopters is a novel and significant contribution to the literature.

The rest of the paper is organized as follows: Sect. 2 presents an iterative algorithm used in this paper to conduct information cascade in a network for a given set of initial adopters and threshold fraction of adopted neighbors. Section 3 presents the proposed binary search algorithm to determine the f_{IA}^{\min} value for a network for a given threshold fraction (q) of adopted neighbors. Section 4 first presents the procedure used to build the step function distribution for q versus f_{IA}^{\min} as well as a binary search algorithm (that makes use of the algorithm of Sect. 3) to determine the q_{step} value and the corresponding f_{IA}^{\min} and $\overline{f_{IA}^{\min}}$ values for a network. Section 4 also explains how the " $1 - q_{step}$ " value (referred to as the Cascade Blocking Index, CBI) for a network could be related to the intra-cluster density of the blocking cluster of the network as well as provides a qualitative analysis of the significance of CBI from the standpoints of information cascade and infection spread. Section 5 first introduces a suite of 40 real-world networks analyzed in this research, presents the $(q_{step}, \text{CBI}, f_{IA}^{\min}, \overline{f_{IA}^{\min}})$ values for these networks with respect to the DEG and BWC centrality metrics and discusses the results from an information cascade standpoint as well as from an infection spread standpoint. Section 5 also runs the well-known Louvain community detection algorithm (Blondel et al. 2008) on the 40 real-world networks, evaluates the intra-cluster densities of the clusters/communities determined for each of these networks, presents a visual comparison of these values with the CBI values observed for these networks as well as analyzes the distribution of the initial adopter nodes in the Louvain clusters of the real-world networks and its impact on the q_{step} , f_{IA}^{\min} and $\overline{f_{IA}^{\min}}$ measures. Finally, Sect. 5 demonstrates the scalability of the binary search algorithms proposed in Sects. 2–4 by measuring and modeling the computation times to determine the q_{step} , f_{IA}^{\min} and $\overline{f_{IA}^{\min}}$ measures for the real-world networks. Section 6 discusses related work in the literature. Section 7 concludes the paper and outlines plans for future work. Throughout the paper, the terms 'node' and 'vertex', 'edge' and 'link', 'network' and 'graph', 'cluster' and 'community', 'metric' and 'measure' are used interchangeably. They mean the same.

Iterative algorithm for information cascade

We use an iterative algorithm (refer to Algorithm 1 for the pseudo code) to conduct information cascade in a network for a given set of initial adopters (IA) and threshold fraction (q) of adopted neighbors. The algorithm runs until all the nodes in the network become part of the set IA (i.e., adopted the unanimous decision, leading to a complete information cascade) or no new node adopted the decision in the latest iteration (tracked through the boolean variable *CascadeProgress* and the set *LatestAdoptedVertices* that are reset to false and \emptyset respectively at the beginning of each iteration). In any iteration, we go through the vertices that have not yet adopted the decision and determine the fraction (f) of their neighbors that have adopted the decision. For any such node u (that has not yet adopted) whose $f \geq q$ (i.e., the fraction of adopted neighbors of the node is greater than or equal to the threshold fraction of adopted neighbors), we include node u to the set *LatestAdoptedVertices*. If *CascadeProgress* stays false at the end of an iteration, it implies no new node adopted the decision during the iteration and the algorithm ends prematurely with the information cascade

declared to be *not complete*. If *CascadeProgress* is true at the end of an iteration, it implies that at least one node has adopted the decision during the iteration and all such vertices in the set *LatestAdoptedVertices* are appended to the set *IA*. Figure 1 illustrates the working of the algorithm.

The algorithm would need to be run for at most $|V| - |IA|$ iterations (a scenario in which just one vertex per iteration adopts the decision) and in each iteration, we go through at most $|E|$ edges of the graph and determine the fraction of adopted neighbors for each vertex in the set $V - IA$. Thus, the overall time complexity of the iterative algorithm is $O(|V||E|)$, simply written as $O(VE)$. As seen in the example of Fig. 1 and the analysis of the real-world networks in Sect. 5, the number of iterations needed to accomplish complete information cascade need not be as large as $|V| - |IA|$ iterations. In Fig. 1, we accomplish complete information cascade in just 2 iterations (wherein $|V| - |IA| = 6$).

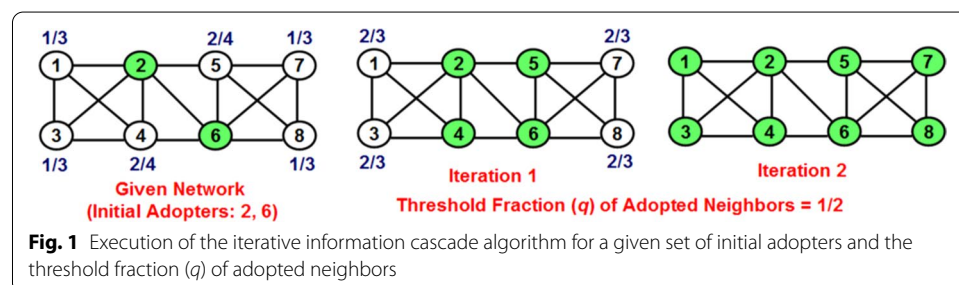
Input: Graph $G = (V, E)$; Set of Initial Adopters, *IA*; Threshold Fraction of Adopted Neighbors, q
Auxiliary Variables: boolean *CascadeProgress*, Set *LatestAdoptedVertices*, int *NumAdoptedNeighbors*
Begin Algorithm *Iterative Information Cascade*

```

while (|IA| < |V|) do
    CascadeProgress = false
    LatestAdoptedVertices =  $\phi$ 
    for vertex  $u \in V - IA$  do
         $N_u$  = Neighbors of vertex  $u$ 
        NumAdoptedNeighbors = 0
        for vertex  $v \in N_u$  do
            if ( $v \in IA$ ) then
                NumAdoptedNeighbors = NumAdoptedNeighbors + 1
            end if
        end for
        Fraction  $f$  of adopted neighbors = NumAdoptedNeighbors /  $|N_u|$ 
        if ( $f \geq q$ ) then
            LatestAdoptedVertices = LatestAdoptedVertices  $\cup \{u\}$ 
            CascadeProgress = true
        end if
    end for
    if (CascadeProgress == false) then
        return "Cascade Not Complete"
    else
        IA = IA  $\cup$  LatestAdoptedVertices
    end if
end while
return "Cascade Complete"

```

End Algorithm *Iterative Information Cascade*



Binary search algorithm to determine the minimum number of initial adopters for a threshold fraction of adopted neighbors

In this section, we present our proposed binary search algorithm (refer to Algorithm 2 for the pseudo code) to determine the minimum number of initial adopters (IA_{min}) needed to operate a network with a threshold fraction (q) of adopted neighbors to accomplish complete information cascade. The minimum fraction of nodes as initial adopters (f_{IA}^{min}) for a network is the minimum number of initial adopters (IA_{min}) determined by the binary search algorithm divided by the number of nodes (N) in the network. The nodes constituting the set of initial adopters are chosen based on a centrality metric (DEG or BWC). As nodes with larger DEG or BWC values were observed to be effective in speeding up the cascade process (Jalili and Perc 2017), in each iteration of the binary search algorithm, we identify the top nodes (nodes having the larger centrality metric values) that constitute the required number of nodes needed as initial adopters (corresponds to the Middle Index value, as explained below and in the pseudo code: Algorithm 2) for the particular iteration. Any ties in choosing the nodes based on the centrality metric are broken arbitrarily. We observe this tie-breaking policy to not result in any significant difference in the results: (q_{step} , f_{IA}^{min} , $\overline{f_{IA}^{min}}$) values of a network across multiple runs of the algorithm in Sect. 5.

Binary search (Cormen et al. 2009) is a classical divide and conquer algorithm design strategy of logarithmic time complexity such that the search space (spanning from a Left Index to the Right Index) gets reduced by half in each iteration and the algorithm stops when the difference between the Right Index and Left Index gets below or equals a termination threshold. The search space for IA_{min} is the range $(0, \dots, N]$, where $N=|V|$, the number of nodes in the network. We maintain the following invariant throughout the execution of the algorithm for a given q : the minimum value for the number of initial adopters needed to accomplish complete information cascade is in the range: (Left Index ... Right Index]. That is, the information cascade will *not* be complete if the number of nodes used as initial adopters corresponds to the Left Index (initially set to 0) and the information cascade will be complete if the number of nodes used as initial adopters corresponds to the Right Index (initially set to N).

In the beginning of each iteration of the binary search algorithm, we determine the Middle Index as the average of the Left Index and Right Index. We then build a set IA of initial adopters (based on a particular centrality metric C) such that the number of nodes constituting the set IA among the nodes in the network equals the value of the Middle Index. We then conduct information cascade on the network (by running the *Iterative Information Cascade* algorithm of Sect. 2) with the set IA of initial adopters and the threshold fraction (q) of adopted neighbors. If the *Iterative Information Cascade* algorithm run for the set IA corresponding to the Middle Index and q is complete, it implies the cascade will also be complete for $|IA|$ values (i.e., the number of initial adopters) greater than the Middle Index. Hence, in such a case, we move the Right Index to the left (as part of the binary search strategy of halving the search space in each iteration) and set the Right Index = Middle Index. If the *Iterative Information Cascade* algorithm run for the set IA corresponding to the Middle Index and q is *not* complete, it implies the cascade will also not be complete for $|IA|$ values lower than the Middle Index. Hence, in such a case, we move the Left Index to

the right and set the Left Index = Middle Index. We stop the iterations when the difference between the Right Index and Left Index equals 1 and we consider the latest value of the Right Index as the minimum number of initial adopters (IA_{min}) needed to operate the network with a threshold fraction (q) of adopted neighbors leading to a complete information cascade. We return $f_{IA}^{min} = |IA_{min}|/N$ as the minimum fraction of initiator nodes needed to operate the network with a threshold q .

Inputs: Graph $G = (V, E)$; Threshold Fraction (q) of Adopted Neighbors;
Centrality metric (C) values of the vertices

Output: Minimum Fraction of Nodes as Initial Adopters, f_{IA}^{min}

Auxiliary Variables: Left Index = 0; Right Index = $N = |V|$; IA_{min} ; Middle Index

Begin Algorithm *Minimum_Initial_Adopters*

while (Right Index - Left Index > 1) **do**

 Middle Index = (Left Index + Right Index) / 2

 Find the set IA of initial adopters using centrality metric C such that $|IA|/N = \text{Middle Index}$

if (*Iterative Information Cascade* on G using the set IA and threshold fraction q is complete) **then**

 Right Index = Middle Index

else

 Left Index = Middle Index

end if

end while

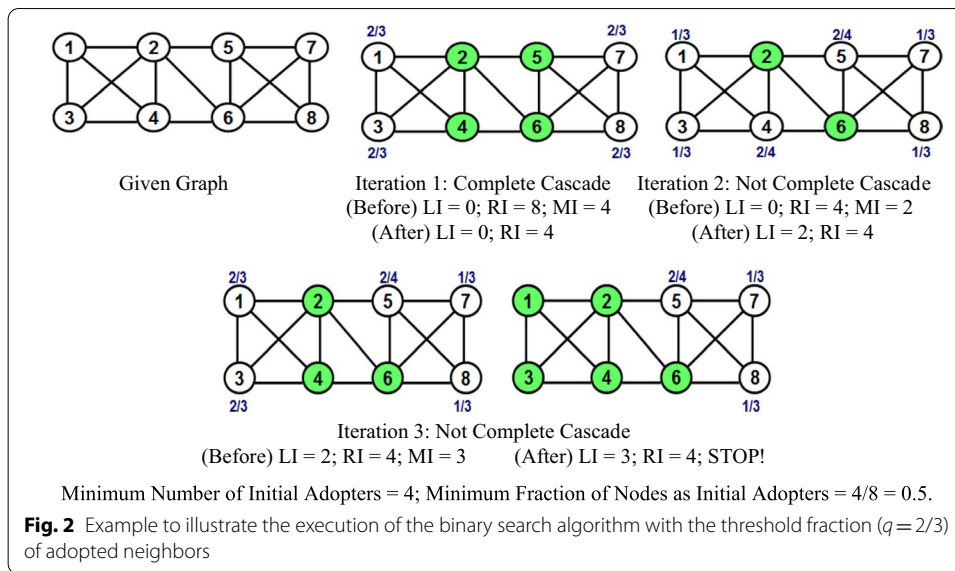
$IA_{min} = \text{Right Index}$

return $f_{IA}^{min} = IA_{min} / N$

End Algorithm *Minimum_Initial_Adopters*

The number of iterations of the proposed binary search algorithm is $\log_2(|V|)$, where $|V|$ is the number of nodes in the network and is the size of the search space at the beginning of the first iteration. The overall time complexity of the algorithm depends on the number of iterations of the algorithm and the time complexity of the Iterative Information Cascade algorithm run for each iteration. The overall time complexity of the Iterative Information Cascade algorithm is $O(|V|*|E|)$. The overall time complexity of the proposed *Minimum_Initial_Adopters* binary search algorithm is then $O(|V|*|E|*\log_2(|V|))$ and is simply denoted as $O(EV\log V)$.

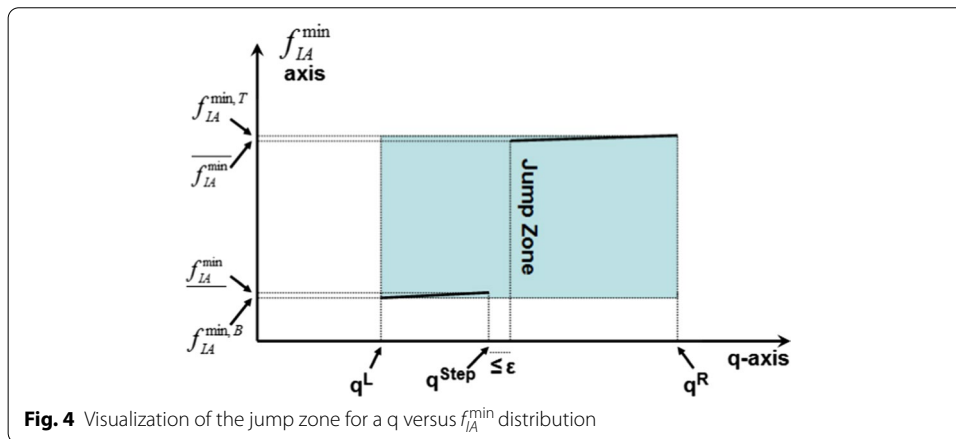
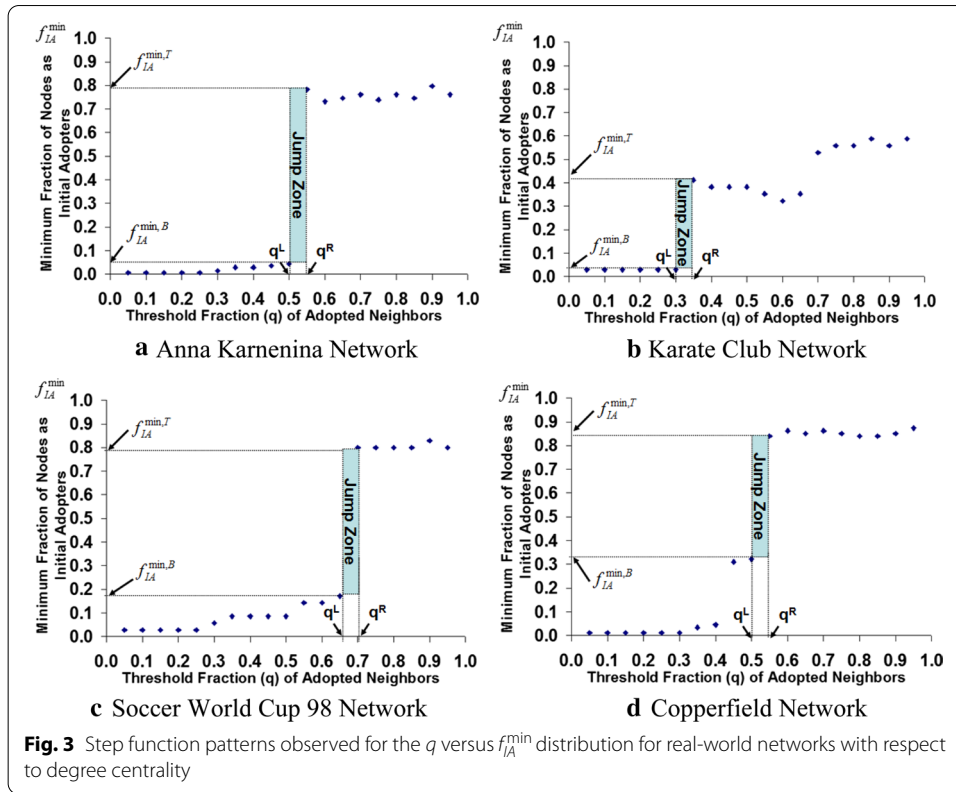
Figure 2 presents an example to illustrate the working of the *Minimum_Initial_Adopters* binary search algorithm. The example graph has 8 vertices and we seek to find the minimum number/minimum fraction of initial adopters needed to operate the network at a threshold fraction ($q = 2/3$) of adopted neighbors to lead to a complete information cascade. Assume the centrality metric used is the degree centrality (DEG) metric, which is the number of neighbors for a node. The initial values of the Left Index (LI) and Right Index (RI) are 0 and 8 respectively. In the first iteration, the Middle Index (MI) value = $(0 + 8)/2 = 4$. The top four vertices with the largest DEG centrality are: 2, 6, 4 and 5. With these four vertices as initial adopters, we see the fraction of adopted neighbors for the other four vertices 1, 3, 7 and 8 to be $2/3$ each, which is equal to q . Hence, all the four vertices 1, 3, 7 and 8 adopt the decision of their neighbors, leading to a complete information cascade. Hence, we move the Right Index to the left and set the Right Index = Middle Index = 4.



In the second iteration, the value for the Middle Index is $(0 + 4)/2 = 2$. The top two vertices with the largest DEG centrality are 2 and 6. With these two initial adopters, the fraction of adopted neighbors for vertices 4 and 5 are $2/4 = 1/2$ each and for the other four vertices 1, 3, 7 and 8 are $1/3$ each. None of these fractions of adopted neighbors are greater than or equal to $q = 2/3$. Hence, the information cascade is considered to be *not* complete and we move the Left Index to the right by setting Left Index = Middle Index = 2. In the third iteration, the value for the Middle Index is $(2 + 4)/2 = 3$. The top three vertices with the largest DEG centrality are 2, 6 and 4 (we can pick either 4 or 5; we break the tie arbitrarily in favor of 4). With these three initial adopters, vertices 1 and 3 will have $2/3$ rd of adopted neighbors, which is equal to q and hence will get added to the set of adopters. However, the fractions of adopted neighbors for vertices 5, 7 and 8 will not be affected by this and they will remain to be less than $2/3$ as shown in Fig. 2. Hence, the information cascade has to be declared *not complete* and we set the Left Index = Middle Index = 3. At the end of the third iteration, the difference between the Right Index and Left Index has reached 1 and we exit the algorithm. The minimum number of initial adopters needed is 4 (the latest value of the Right Index) to operate the network at a threshold fraction $q = 2/3$ of adopted neighbors. The corresponding minimum fraction of nodes as initial adopters is $4/8 = 0.5$.

Analysis of the relationship between the threshold fraction of adopted neighbors versus the minimum fraction of nodes as initial adopters and the intra-cluster density of blocking cluster

In this section, we first empirically analyze the relationship between the threshold fraction (q) of adopted neighbors for a complete information cascade versus the minimum fraction of nodes to be used as initial adopters (f_{IA}^{\min}). We then analyze the relationship between the intra-cluster density of the blocking cluster of a network and the above two parameters (q and f_{IA}^{\min}).



Analysis of the relationship: q versus f_{LA}^{\min}

For each of the 40 complex real-world networks analyzed in this paper and a centrality metric (DEG or BWC), we ran the binary search algorithm of Sect. 3 for q values ranging from 0.05 to 0.95, in increments of 0.05, and recorded the f_{LA}^{\min} values. We plot the q versus f_{LA}^{\min} values in a two-dimensional coordinate system and observed the distribution to exhibit a step function pattern for 37 of the 40 complex real-world networks considered and the two centrality metrics. As we increase the q value, there appear one or more spikes in the f_{LA}^{\min} values. We refer to the spike that has the "largest" increase in the f_{LA}^{\min} value as the jump zone (see Fig. 4), which is of width 0.05 (spanned by the q values q^L

and q^R to the left and right) and height $f_{IA}^{\min,T} - f_{IA}^{\min,B}$, wherein $f_{IA}^{\min,T}$ and $f_{IA}^{\min,B}$ are the f_{IA}^{\min} values when $q=q^R$ and $q=q^L$ respectively (i.e., at the top and bottom of the jump zone corresponding to the f_{IA}^{\min} axis).

Figure 3 shows the typical step function patterns that we noticed for the real-world networks analyzed in this research: (a) There is only one jump zone and the f_{IA}^{\min} values appear not to change much before and after the jump zone. (b) There are two or more spikes and the first spike in the f_{IA}^{\min} value corresponds to the jump zone. (c) There is only one spike, but the f_{IA}^{\min} value gradually increases (with increase in q) before and/or after the jump zone. (d) There are two or more spikes and the second spike in the f_{IA}^{\min} value corresponds to the jump zone. The Copperfield network exemplifies that the first spike in the f_{IA}^{\min} value (unlike the Karate network) need not always correspond to the jump zone.

Our next step in the analysis is to zoom on the jump zone and determine the exact value of q (referred to as q_{step}) and the corresponding f_{IA}^{\min} values (referred to as $\underline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$) such that $f_{IA}^{\min,B} \sim \underline{f_{IA}^{\min}} < \overline{f_{IA}^{\min}} < f_{IA}^{\min,T}$ and $f_{IA}^{\min,B} < \underline{f_{IA}^{\min}} \sim f_{IA}^{\min,T}$. Note that $\underline{f_{IA}^{\min}}$ is the f_{IA}^{\min} value when $q=q_{step}$ and $\overline{f_{IA}^{\min}}$ is the f_{IA}^{\min} value at the next measurable value of q beyond q_{step} . Overall, $f_{IA}^{\min,B} \leq \underline{f_{IA}^{\min}} < \overline{f_{IA}^{\min}} \leq f_{IA}^{\min,T}$ in the f_{IA}^{\min} -axis and $q^L \leq q_{step} < q^R$ in the q -axis. Figure 4 presents a visualization of the variables q_{step} , $\underline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$ for a jump zone of width $q^R - q^L$ and height $f_{IA}^{\min,T} - f_{IA}^{\min,B}$. The absolute value of the difference between q_{step} and the next measurable value of q is less than or equal to ε ($\varepsilon=0.001$ in this paper), a parameter referred to as the termination threshold of the binary search algorithm that we will now describe to determine the exact value of q_{step} .

The binary search algorithm (referred to as *Jump Zone Analyzer*; pseudo code in Algorithm 3) to determine the exact value of q_{step} has a search space of $[q^L, \dots, q^R]$. Appropriately, the algorithm starts with the Left Index set to q^L and the Right Index set to q^R whose corresponding f_{IA}^{\min} values are $f_{IA}^{\min,B}$ and $f_{IA}^{\min,T}$ respectively. The following invariant is satisfied throughout the execution of the algorithm: The Left Index always takes up a q value for which the corresponding f_{IA}^{\min} value (indicated as $f_{IA}^{\min,LeftIndex}$) is much closer to $f_{IA}^{\min,B}$ and much lower than $f_{IA}^{\min,T}$. That is, $|f_{IA}^{\min,LeftIndex} - f_{IA}^{\min,B}| < |f_{IA}^{\min,T} - f_{IA}^{\min,LeftIndex}|$. Similarly, the Right Index always takes up a q value for which the corresponding f_{IA}^{\min} value (indicated as $f_{IA}^{\min,RightIndex}$) is much closer to $f_{IA}^{\min,T}$ and much greater than $f_{IA}^{\min,B}$. That is $|f_{IA}^{\min,T} - f_{IA}^{\min,RightIndex}| < |f_{IA}^{\min,RightIndex} - f_{IA}^{\min,B}|$.

Inputs: Graph $G = (V, E)$; Centrality metric (C) values of the vertices; Termination Threshold (ϵ); Threshold Fractions q^L and q^R of adopted neighbors, corresponding to the jump zone.

Outputs: Threshold Fraction q_{step} , and the f_{IA}^{\min} values $\overline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$.

Auxiliary Variables: Left Index; Right Index; Middle Index; $f_{IA}^{\min, LeftIndex}$, $f_{IA}^{\min, RightIndex}$, $f_{IA}^{\min, MiddleIndex}$

Initialization: Left Index = q^L ; Right Index = q^R
 $f_{IA}^{\min, LeftIndex}$ = Run algorithm *Minimum_Initial_Adopters* (G ; q^L ; C)
 $f_{IA}^{\min, RightIndex}$ = Run algorithm *Minimum_Initial_Adopters* (G ; q^R ; C)

Begin Algorithm Jump Zone Analyzer

while ($|Right\ Index - Left\ Index| \geq \epsilon$) **do**
 Middle Index = (Left Index + Right Index) / 2
 $f_{IA}^{\min, MiddleIndex}$ = Run algorithm *Minimum_Initial_Adopters* (G ; Middle Index; C)

if ($|f_{IA}^{\min, MiddleIndex} - f_{IA}^{\min, LeftIndex}| \leq |f_{IA}^{\min, RightIndex} - f_{IA}^{\min, MiddleIndex}|$) **then**
 Left Index = Middle Index
 else
 Right Index = Middle Index
 end if
end while

return $q_{step} = Left\ Index$; $\overline{f_{IA}^{\min}} = f_{IA}^{\min, LeftIndex}$; $\overline{f_{IA}^{\min}} = f_{IA}^{\min, RightIndex}$

End Algorithm Jump Zone Analyzer

In the beginning of each iteration of the algorithm, we determine the Middle Index to be the average of the Left Index and Right Index and use the binary search algorithm (Algorithm *Minimum_Initial_Adopters*) of Sect. 3 to determine the f_{IA}^{\min} value (indicated as $f_{IA}^{\min, MiddleIndex}$) when the Middle Index value is used as the threshold fraction (q) of adopted neighbors. If $|f_{IA}^{\min, MiddleIndex} - f_{IA}^{\min, LeftIndex}| \leq |f_{IA}^{\min, RightIndex} - f_{IA}^{\min, MiddleIndex}|$, it implies the $f_{IA}^{\min, MiddleIndex}$ value is much closer to $f_{IA}^{\min, LeftIndex}$ and is lower than $f_{IA}^{\min, RightIndex}$: This will be the case for any q value ranging from the Left Index to the Middle Index. Hence, we move the Left Index to the right and set the Left Index = Middle Index. If $|f_{IA}^{\min, RightIndex} - f_{IA}^{\min, MiddleIndex}| < |f_{IA}^{\min, MiddleIndex} - f_{IA}^{\min, LeftIndex}|$, it implies the $f_{IA}^{\min, MiddleIndex}$ value is much closer to $f_{IA}^{\min, RightIndex}$ and is greater than $f_{IA}^{\min, LeftIndex}$: This will be the case for any q value ranging from the Middle Index to the Right Index. Hence, we move the Right Index to the left and set the Right Index = Middle Index. We continue the iterations as long as the absolute difference between the Right Index and Left Index stays greater than or equal to a termination threshold (ϵ). We stop the algorithm when the absolute difference between the Right Index and Left Index becomes less than ϵ and declare the latest values of the Left Index, $f_{IA}^{\min, LeftIndex}$ and $f_{IA}^{\min, RightIndex}$ as q_{step} , $\overline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$ respectively.

With a search space of width 0.05 and termination threshold $\epsilon = 0.001$, the number of iterations of the *Jump Zone Analyzer* algorithm is $\log_2(0.05/0.001) \sim 6$. As we run the binary search algorithm *Minimum_Initial_Adopters* of Sect. 3 in each of the six iterations, the overall time complexity of the binary search algorithm *Jump Zone Analyzer* to determine the q_{step} value is the same as the overall time complexity of the binary search algorithm *Minimum_Initial_Adopters*.

Table 1 illustrates the execution of the *Jump Zone Analyzer* binary search algorithm for the jump zone of the Karate Club network of Fig. 3b whose q^L and q^R values

Table 1 Execution of the *Jump Zone Analyzer* binary search algorithm on the jump zone for the karate club network

It #	LI	RI	$ RI-LI , \geq 0.001, ?$	MI	$f_{IA}^{\min,LI}$	$f_{IA}^{\min,RI}$	$f_{IA}^{\min,MI}$	$\left(\frac{ f_{IA}^{\min,MI} - f_{IA}^{\min,LI} }{ f_{IA}^{\min,RI} - f_{IA}^{\min,MI} } \leq \right), ?$	Change
1	0.30	0.35	0.05, Yes	0.325	0.0294	0.4118	0.0294	$0 \leq 0.3824$, Yes	LI = MI
2	0.325	0.35	0.025, Yes	0.3375	0.0294	0.4118	0.3823	$0.3529 \leq 0.0295$, No	RI = MI
3	0.325	0.3375	0.0125, Yes	0.3313	0.0294	0.3823	0.0294	$0 \leq 0.3529$, Yes	LI = MI
4	0.3313	0.3375	0.0062, Yes	0.3344	0.0294	0.3823	0.3823	$0.3529 \leq 0$, No	RI = MI
5	0.3313	0.3344	0.0031, Yes	0.3329	0.0294	0.3823	0.0294	$0 \leq 0.3529$, Yes	LI = MI
6	0.3329	0.3344	0.0015, Yes	0.3337	0.0294	0.3823	0.3529	$0.3235 \leq 0.0294$, No	RI = MI
7	0.3329	0.3337	0.0008, No	STOP!! $q_{step} = 0.3329$; $\underline{f_{IA}^{\min}} = 0.0294$; $\overline{f_{IA}^{\min}} = 0.3823$					

are 0.30 (initial Left Index) and 0.35 (initial Right Index) respectively, and the termination threshold $\varepsilon = 0.001$. For ease of accommodating multiple columns in Table 1, we refer to the Left Index, Right Index and Middle Index using the acronyms LI, RI and MI respectively. Based on the initial LI (q^L) and RI (q^R) values, the initial values for $f_{IA}^{\min,LI}$ ($f_{IA}^{\min,LeftIndex}$) and $f_{IA}^{\min,RI}$ ($f_{IA}^{\min,RightIndex}$) are 0.0294 and 0.4118 respectively. All values are rounded to four decimals for ease of representation in Table 1 (note that there is no rounding of numbers in the executions conducted on real-world networks in Sect. 5). As noticed in Table 1, there are a total of 6 complete iterations and the algorithm stops at the beginning of the 7th iteration (with the difference between RI and LI becoming less than 0.001). The algorithm outputs the latest values of the LI, $f_{IA}^{\min,LI}$ and $f_{IA}^{\min,RI}$ as $q_{step} = 0.3329$, $\underline{f_{IA}^{\min}} = 0.0294$ and $\overline{f_{IA}^{\min}} = 0.3823$ respectively.

Cascade Blocking Index: $1 - q^{step}$ and intra-cluster density of blocking cluster

Clusters are expected to have a larger intra-cluster density and lower inter-cluster density, and such clusters are referred to as *modular* clusters (Easley and Kleinberg 2010; Newman 2010). Modular clusters have the potential to block information cascade from penetrating to nodes inside the cluster. For information cascade to penetrate into a cluster and result in a complete cascade, the bridge nodes of the cluster need to first adopt the unanimous decision: the internal nodes of the cluster would be able to adopt a decision only if at least the threshold fraction of its neighbor nodes (primarily the bridge nodes) have adopted the decision. For a bridge node to adopt a decision, its fraction of adopted neighbors (that are outside the cluster) should be greater than or equal to the threshold fraction (q) of adopted neighbors for the cascade. For a cluster with high intra-cluster density, the bridge nodes of the cluster are also expected to have a larger intra-cluster density: i.e., the fraction of adopted neighbors (outside the cluster) of the bridge nodes is expected to be lower. Hence, for information cascade to penetrate into such a cluster with high intra-cluster density, the threshold fraction (q) of adopted neighbors must be lower. Conversely, the larger the threshold fraction (q) of adopted neighbors with which we can accomplish complete information cascade in a network, lower the intra-cluster density of the clusters in the network.

For a cluster of intra-cluster density ρ , the fraction of adopted neighbors (outside the cluster) possible for a bridge node is at most $1 - \rho$. In order to make the bridge node to adopt the unanimous decision needed for complete information cascade, $1 - \rho \geq q$;

i.e., the intra-cluster density $\rho \leq 1 - q$. If $1 - q < \rho$ (i.e., $\rho > 1 - q$), the cluster will not be penetrable (i.e., the bridge nodes cannot be made to adopt a decision) and will become a blocking cluster. Hence, in order to be able penetrate such blocking clusters of high intra-cluster density, it becomes imperative to include one or more bridge nodes of these blocking clusters as part of the set of initial adopters.

In Sect. 4.1, we noticed that the minimum fraction of nodes as initial adopters needed for complete information cascade increased in a step function pattern from f_{IA}^{\min} to $\overline{f_{IA}^{\min}}$ and q_{step} (corresponding to $\overline{f_{IA}^{\min}}$) is the last measurable value of q beyond which we needed to increase the minimum fraction of nodes as initial adopters from f_{IA}^{\min} to $\overline{f_{IA}^{\min}}$ in order to accomplish complete information cascade. That is, when the intra-cluster densities (ρ values) of the clusters in the network were all less than or equal to $1 - q_{step}$, we were able to accomplish complete information cascade with f_{IA}^{\min} values $\leq \overline{f_{IA}^{\min}}$. Any further increase in q (beyond q_{step}) would make the intra-cluster density of at least one cluster to become greater than $1 - q$ and we will not be able to penetrate such a cluster with $\overline{f_{IA}^{\min}}$ fraction of initial adopters and will need to increase the minimum fraction of initial adopters to $\overline{f_{IA}^{\min}}$, with $\overline{f_{IA}^{\min}}$ appreciably greater than f_{IA}^{\min} for most of the networks. Hence, in order to capture the intra-cluster density of such blocking clusters, we refer to the jump zone (refer Figs. 3, 4) as the bounded area in the q versus f_{IA}^{\min} distribution that encounters the largest increase in the f_{IA}^{\min} values. The value of " $1 - q_{step}$ " for a network could be thus used to assess the intra-cluster density of the blocking cluster(s) for the network. We propose that the $1 - q_{step}$ value for a network be called the Cascade Blocking Index (CBI) of the network (whose values range from 0 to 1), a quantitative measure of the intra-cluster density of the blocking cluster(s) of the network. The larger the CBI value for a network, the larger the intra-cluster density of the blocking cluster(s) of the network and vice-versa.

Cascade Blocking Index: information cascade versus infection spread

It is important to note that the CBI metric can be used to decide on the nature of values for the threshold fraction of adopted neighbors for accomplishing complete information cascade. For networks with a lower CBI value, one could impose a larger value for the threshold fraction of adopted neighbors that are needed for a node to adopt a decision and be able to still accomplish complete information cascade (the bridge nodes of the clusters will have a larger fraction of neighbors that are outside the clusters and if these neighbors adopt a decision, the bridge nodes will also be in a position to adopt the decision). For networks with lower CBI value, it is thus possible to make the nodes to consensually adopt the unanimous decision (needed for accomplishing complete information cascade) only after a majority of its neighbors adopt the same decision (i.e., the nodes need not be in a rush to adopt the decision when only a minority of their neighbors have adopted). On the other hand, for networks with a larger CBI value, one would have to operate at relatively lower values for the threshold fraction of adopted neighbors to accomplish complete information cascade. That is, for networks with a larger CBI value, nodes (especially, the bridge nodes) may be forced to adopt the unanimous decision needed for complete information cascade even if a majority of its neighbors have not yet decided by then.

From an epidemic standpoint, we do not want a virus/disease to be able to penetrate the clusters of a network and spread. In this context, if the bridge nodes are say vaccinated (to be immune to the disease; i.e., operating at q values closer to 1), unless all the neighbors become infected, the bridge nodes are not vulnerable to get infected and the virus/disease cannot penetrate through the clusters. For clusters with lower intra-cluster density (i.e., lower CBI values), the bridge nodes are expected to have a larger fraction of neighbors who are outside the cluster. If the bridge nodes are not immune to the disease (i.e., the q value at which the network can be operated is lower) and the network has a lower CBI value, it is possible for the bridge nodes to get infected even if a lower fraction of the outside neighbors get infected and the disease can easily penetrate through the clusters. Thus, if the network has a lower CBI value, it is essential to immunize the bridge nodes of the clusters. On the other hand, if a network has a larger CBI value (i.e., a lower fraction of the neighbors for the bridge nodes are outside the clusters) and the bridge nodes are immune to the disease (i.e., the q value at which the network can be operated could be higher, closer to 1), it will be difficult for the infection to penetrate through a cluster from outside through these bridge nodes.

Quantitative analysis of the real-world networks

We analyzed a suite of 40 real-world networks of diverse domains. Table 2 presents a listing of the 40 real-world networks along with the number of nodes and edges, their references, domain and the three-character code used to refer to these networks later in the paper. The networks analyzed are spread over several domains (the numbers inside the parenthesis indicate the number of networks analyzed from these domains) like co-appearance networks (7), biological networks (4), collaboration networks (3), literature networks (2), employee networks (3), transportation networks (2), game network (1), geographical network (1), citation network (1), social networks (13) and web networks (3). All the networks are considered as undirected graphs and are connected (i.e., the vertices in a graph are reachable to each other and exist as a single component). Table 3 (obtained by running the Jump Analyzer binary search algorithm of Sect. 4) presents the $\text{CBI} = 1 - q_{\text{step}}$, $\overline{f_{IA}^{\min}}$ and f_{IA}^{\min} values for the 40 real-world networks with respect to the degree (DEG) and betweenness (BWC) centrality metrics (for choosing the initial adopters).

Figure 5 plots the difference $\overline{f_{IA}^{\min}} - f_{IA}^{\min}$ in the f_{IA}^{\min} values in the vicinity of q_{step} values in the jump zones observed for the 40 real-world networks. We observe an appreciable difference (of 0.14 or more) in the $\overline{f_{IA}^{\min}} - f_{IA}^{\min}$ values (with respect to both DEG and BWC) for 37 of the 40 real-world networks. The median of the $\overline{f_{IA}^{\min}} - f_{IA}^{\min}$ values is observed to be 0.44 for DEG-based analysis and 0.32 for BWC-based analysis. We observe the difference to go as large as 0.70, justifying our association of the increase in the f_{IA}^{\min} values in the vicinity of q_{step} to the cascade capacity of the blocking cluster(s) of the networks.

Analysis of the CBI values

Figure 6 displays the CBI values for the 40 real-world networks with respect to DEG and BWC. The median for both the CBI(DEG) and CBI(BWC) values is 0.50. We observe the CBI(DEG) values to be 0.50 or more for 30 of the 40 networks (i.e., for 75% of the

Table 2 Real-world networks used for the analysis

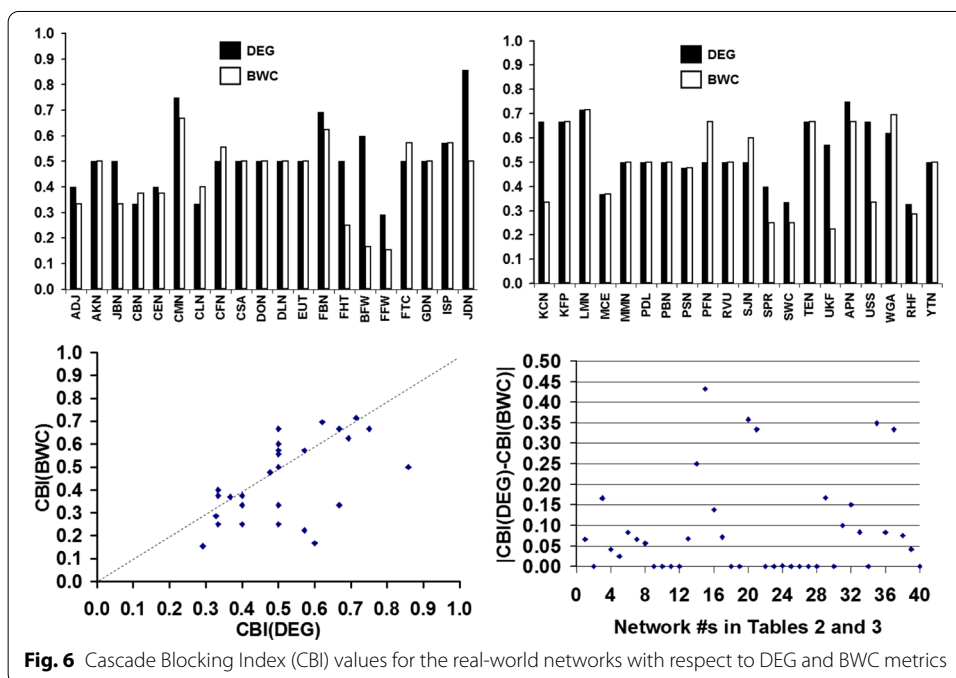
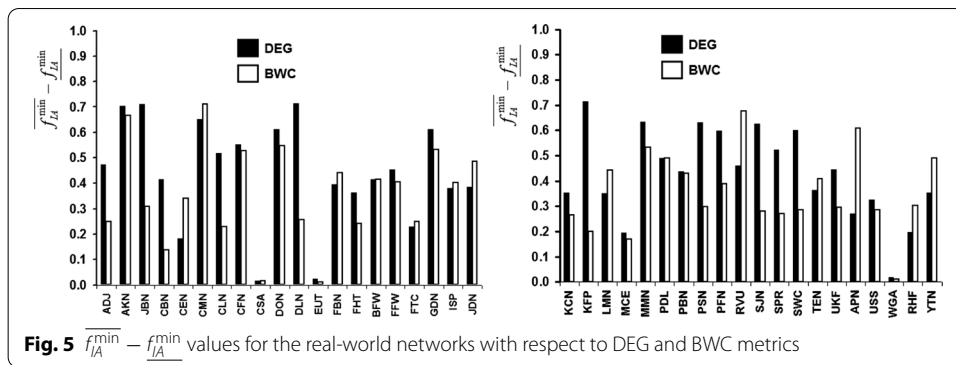
#	Network name	# Nodes	# Edges	References	Domain	Code
1	Word adjacency network	112	425	Newman (2006)	Co-appearance net	ADJ
2	Anna Karenina network	138	493	Knuth (1993)	Co-appearance net	AKN
3	Jazz band network	198	2742	Geiser and Danon (2003)	Co-appearance net	JBN
4	Cat brain network	65	730	Scannell et al. (1995)	Biological network	CBN
5	C. Elegans neural network	297	2148	White et al. (1986)	Biological network	CEN
6	C. Elegans metabolic network	453	2025	Duch and Arenas (2005)	Biological network	CMN
7	Centrality literature network	118	613	Hummon et al. (1990)	Literature network	CLN
8	Copperfield network	87	406	Knuth (1993)	Co-appearance net	CFN
9	CS Department Aarhus Network	61	219	Magnani et al. (1303)	Employee network	CSA
10	Dolphin network	62	159	Lusseau et al. (2003)	Social network	DON
11	Dutch literature 1976 network	35	80	Nooy (1999)	Literature network	DLN
12	EU air transportation network	418	1999	Cadrillo et al. (1344)	Transportation net	EUT
13	Facebook network	324	2218	Blagus et al. (2012)	Social network	FBN
14	Friendship ties in a hi-tech firm	33	147	Krackhardt (1999)	Employee network	FHT
15	Flensburg fjord biological food web	180	1577	Zander (2011)	Web network	BFW
16	Florida food web carbon exchange	128	2106	Ulanowicz and Donald (2005)	Web network	FFW
17	Flying teams cadet network	48	170	Mareno (1960)	Social network	FTC
18	GD'96 network	180	229	Batagelj et al. (2006)	Citation network	GDN
19	Infectious socio-patterns network	309	1924	Isella et al. (2011)	Social network	ISP
20	Java dependency network	1538	8032	Batagelj et al. (2006)	Collaboration net	JDN
21	Karate club network	34	78	Zachary (1977)	Social network	KCN
22	Korea family planning network	35	84	Rogers and Kincaid (1980)	Social network	KFP
23	Les Miserables network	77	254	Knuth (1993)	Co-appearance net	LMN
24	Manufacturing company employee net	77	2326	Cross et al. (2004)	Employee network	MCE
25	ModMath network	30	61	Batagelj et al. (2006)	Social network	MMN
26	Perl developers network	839	2111	Failed (2009)	Collaboration net	PDL
27	US politics books network	105	441	Krebs (2003)	Co-appearance net	PBN
28	Primary school contact network	238	5539	Gemmetto et al. (2014)	Social network	PSN
29	Prison friendship network	67	182	MacRae (1960)	Social network	PFN
30	RVU email network	1133	10,903	Guimera et al. (2003)	Collaboration net	RVU
31	San Juan Sur family network	75	155	Loomis et al. (1953)	Social network	SJN
32	Senator press release network	92	477	Grimmer (2010)	Co-appearance net	SPR
33	Soccer World Cup 1998 Network	35	118	Batagelj et al. (2006)	Game network	SWC
34	Taro Exchange Network	22	39	Schwimmer (1973)	Social network	TEN
35	UK Faculty Friendship Network	81	577	Nepusz et al. (2008)	Social network	UKF
36	US Airports 1997 Network	332	2126	Batagelj et al. (2006)	Transportation net	APN
37	US States Network	49	107	Meghanathan (2017a)	Geographical net	USS
38	Web Graph of Amazon Pages	2879	3886	Subelj and Bajec (2012)	Web network	WGA

Table 2 (continued)

#	Network name	# Nodes	# Edges	References	Domain	Code
39	Residence Hall Friendship Network	217	1839	Freeman et al. (1998)	Social network	RHF
40	Yeast Transcription Network	4441	12,873	Bhardwaj et al. (2010)	Biological network	YTN

Table 3 Results obtained for the Real-World Networks

Results with respect to DEG centrality					Results with respect to BWC centrality				
#	Net. code	CBI: $1 - q_{step}$	f_{IA}^{min}	$\overline{f_{IA}^{min}}$	#	Net. code	CBI: $1 - q_{step}$	f_{IA}^{min}	$\overline{f_{IA}^{min}}$
1	ADJ	0.4000	0.1607	0.6339	1	ADJ	0.3336	0.2411	0.4911
2	AKN	0.5000	0.0435	0.7464	2	AKN	0.5000	0.0435	0.7101
3	JBN	0.5000	0.2273	0.9394	3	JBN	0.3336	0.4040	0.7121
4	CBN	0.3336	0.3846	0.8000	4	CBN	0.3750	0.3077	0.4462
5	CEN	0.4000	0.1549	0.3367	5	CEN	0.3750	0.1650	0.5051
6	CMN	0.7500	0.0022	0.6534	6	CMN	0.6672	0.1656	0.8764
7	CLN	0.3336	0.2627	0.7797	7	CLN	0.4000	0.1441	0.3729
8	CFN	0.5000	0.3103	0.8621	8	CFN	0.5563	0.0575	0.5862
9	CSA	0.5000	0.0328	0.0492	9	CSA	0.5000	0.0328	0.0492
10	DON	0.5000	0.1774	0.7903	10	DON	0.5000	0.2097	0.7581
11	DLN	0.5000	0.0857	0.8000	11	DLN	0.5000	0.1714	0.4286
12	EUT	0.5000	0.0311	0.0550	12	EUT	0.5000	0.0383	0.0502
13	FBN	0.6930	0.0370	0.4321	13	FBN	0.6250	0.1420	0.5833
14	FHT	0.5000	0.1212	0.4848	14	FHT	0.2500	0.3939	0.6364
15	BFW	0.6000	0.0444	0.4611	15	BFW	0.1672	0.2667	0.6833
16	FFW	0.2922	0.3125	0.7656	16	FFW	0.1539	0.4531	0.8594
17	FTC	0.5000	0.2708	0.5000	17	FTC	0.5719	0.0833	0.3333
18	GDN	0.5000	0.0444	0.6556	18	GDN	0.5000	0.0444	0.5778
19	ISP	0.5719	0.3463	0.7282	19	ISP	0.5719	0.0874	0.4919
20	JDN	0.8578	0.0007	0.3875	20	JDN	0.5000	0.2341	0.7191
21	KCN	0.6672	0.0294	0.3824	21	KCN	0.3336	0.3529	0.6176
22	KFP	0.6672	0.0571	0.7714	22	KFP	0.6672	0.1143	0.3143
23	LMN	0.7148	0.0130	0.3636	23	LMN	0.7148	0.0130	0.4545
24	MCE	0.3672	0.4545	0.6494	24	MCE	0.3688	0.3766	0.5455
25	MMN	0.5000	0.2667	0.9000	25	MMN	0.5000	0.2000	0.7333
26	PDL	0.5000	0.0083	0.4994	26	PDL	0.5000	0.0083	0.4994
27	PBN	0.5000	0.0952	0.5333	27	PBN	0.5000	0.1714	0.6000
28	PSN	0.4766	0.1933	0.8235	28	PSN	0.4766	0.2101	0.5084
29	PFN	0.5000	0.2836	0.8806	29	PFN	0.6672	0.0299	0.4179
30	RVU	0.5000	0.3875	0.8482	30	RVU	0.5000	0.2365	0.9126
31	SJN	0.5000	0.2400	0.8667	31	SJN	0.6000	0.0800	0.3600
32	SPR	0.4000	0.1739	0.6957	32	SPR	0.2500	0.5217	0.7935
33	SWC	0.3336	0.2000	0.8000	33	SWC	0.2500	0.4857	0.7714
34	TEN	0.6672	0.0455	0.4091	34	TEN	0.6672	0.0455	0.4545
35	UKF	0.5719	0.1111	0.5556	35	UKF	0.2227	0.4691	0.7654
36	APN	0.7500	0.2440	0.5151	36	APN	0.6672	0.0572	0.6657
37	USS	0.6672	0.0408	0.3673	37	USS	0.3336	0.5510	0.8367
38	WGA	0.6203	0.0031	0.0049	38	WGA	0.6961	0.0024	0.0035
39	RHF	0.3281	0.4885	0.6866	39	RHF	0.2859	0.5161	0.8203
40	YTN	0.5000	0.0059	0.3605	40	YTN	0.5000	0.0065	0.4970



networks) and the CBI(BWC) values to be 0.50 or more for 24 of the 40 networks (i.e., for 60% of the networks). Thus, for about 2/3rds of the real-world networks, we need to operate at lower values for the threshold fraction ($1 - \text{CBI}$; $q < 0.50$) of adopted neighbors and the corresponding f_{IA}^{min} fraction of initial adopters to accomplish complete information cascade. When we apply the qualitative discussion of Sect. 4.3 (CBI: Information Cascade vs. Infection Spread) to the results observed in this section, we can say that for about 2/3rds of the 40 real-world networks analyzed in this paper (that have a larger CBI value), a bridge node is more likely to adopt the unanimous decision (needed for complete information cascade) when less than half of its neighbors have adopted the decision. On the other hand, from an epidemic standpoint, 2/3rds of the 40 real-world networks (that have a larger CBI value) are *not* vulnerable for an infection spread if the bridge nodes of the clusters are vaccinated to the related virus/disease.

With regards to the difference in the CBI values with respect to the DEG and BWC metrics, we observe no difference in the CBI(DEG) and CBI(BWC) values for 11 of the

40 real-world networks (i.e., for 28% of the networks) and a difference of more than 0.10 for only 10 of the 40 real-world networks (i.e., for only 25% of the networks). The median difference in the CBI values is 0.049. On the basis of the raw values, we observe the CBI(DEG) values to be numerically larger than the CBI(BWC) values for 20 of the 40 real-world networks (i.e., for 50% of the networks). We can thus conclude the choice of the centrality metric would only have at most a moderate impact on the CBI values for a network. The relatively lower CBI(BWC) values for certain networks could be due to the larger BWC values for the bridge nodes of the clusters and their preferential inclusion in the set of initial adopters by the *Minimum Initial Adopters* algorithm. As a result, for networks with relatively lower CBI(BWC) values, there is a relatively better chance for accomplishing complete information cascade with larger q values (lowering the $1 - q_{step}$ values) when nodes with larger BWC are considered for inclusion to the set of initial adopters.

CBI values versus intra-cluster densities of the clusters

We ran the Louvain community detection algorithm (Blondel et al. 2008) on the 40 real-world networks and determined the intra-cluster densities of the clusters (identified by the Louvain algorithm) in these networks. The composition of the clusters/communities of a network could vary with the community detection algorithm used to determine them. We chose to use the Louvain community detection algorithm as it is a well-known, computationally-less intensive and commonly used algorithm in the literature and in software packages [like Gephi (Gephi 2020)] to determine modular clusters. The Louvain community detection algorithm is a hierarchical community detection algorithm that recursively merges the communities (initially, each node is in its own community) such that the sum of the modularity scores of the communities is maximized. A cluster/community has a larger modularity score if the intra-cluster density of the cluster is significantly larger than the inter-cluster density. Recall, the intra-cluster density for a cluster is computed as the minimum of the intra-cluster densities of the bridge nodes of the cluster, and the intra-cluster density for a node (including a bridge node) in a cluster is the fraction of its neighbor nodes that are within the same cluster. Figure 7 plots the distribution of the intra-cluster densities of the clusters (in yellow colored-smaller circles) in the 40 real-world networks along with the CBI(DEG) and CBI(BWC) values in red and green-colored larger circles respectively. If the CBI(DEG) and CBI(BWC) values for a network are the same, the red and green circles are placed on top of each other.

With regards to DEG or BWC, we observe CBI(DEG) to be a better choice to quantify the intra-cluster densities of the blocking clusters of the network: the CBI(DEG) values appear as an upper bound for the intra-cluster densities for about 32 of the 40 networks (i.e., for 80% of the networks); the CBI(BWC) values appear as an upper bound for the intra-cluster densities for about 27 of the 40 networks (i.e., for slightly more than 2/3rds of the networks). Either way, the largest of the intra-cluster densities of the clusters (these are the blocking clusters that decide the success or failure of information cascade for a given threshold fraction of adopted neighbors) determined for at least 2/3rds of the real-world networks are less than or equal to the CBI values (with respect to both DEG and BWC) reported in our research.

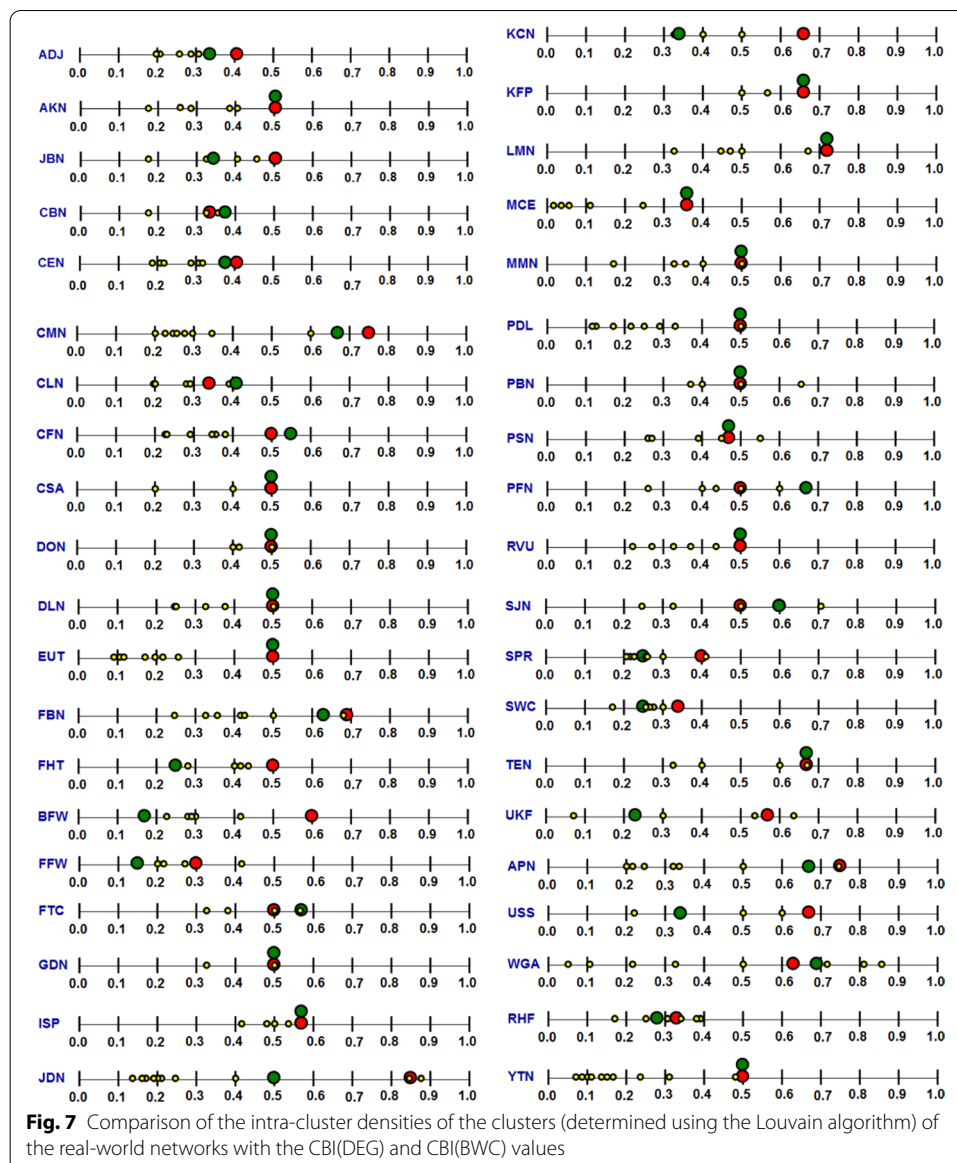
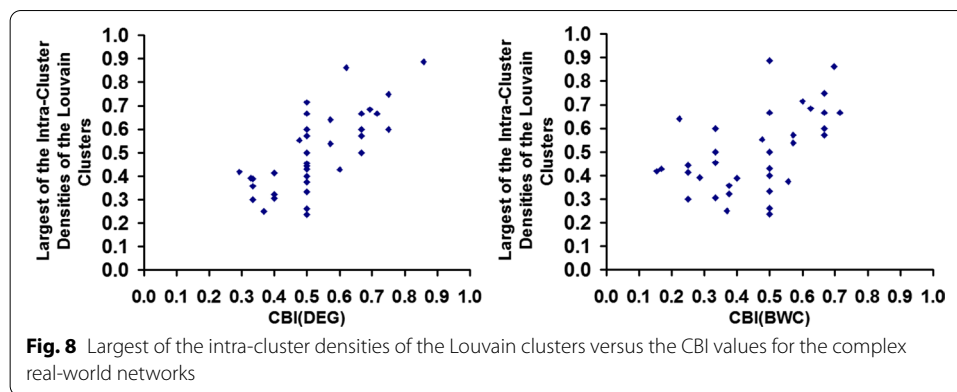


Figure 8 plots the largest of the intra-cluster densities of the Louvain clusters (could be considered as the intra-cluster density of the blocking cluster) of the real-world networks versus the CBI(DEG) and CBI(BWC) values. Table 4 presents the correlation coefficient values observed between these measures: the upper diagonal presents the Pearson's (P) linear regression-based correlation coefficient values and the lower diagonal presents the Spearman's (S) rank-based correlation coefficient values. We observe a strong positive correlation (correlation coefficient values > 0.7) between the CBI(DEG) values for the 40 real-world networks and the largest of the intra-cluster densities of the Louvain clusters in these networks. A strong positive correlation also implies that we could predict the intra-cluster density of the blocking cluster of a network using its CBI value as well as the ranking of the networks based on their CBI

**Table 4** Pearson's (P) and Spearman's (S) correlation coefficient values

	Largest of the intra-cluster densities of the Louvain clusters	CBI(DEG)	CBI(BWC)
Largest of the intra-cluster densities of the Louvain clusters	1.0	0.7250 (P)	0.5129 (P)
CBI(DEG)	0.7181 (S)	1.0	0.5411 (P)
CBI(BWC)	0.5315 (S)	0.5572 (S)	1.0

values would be almost the same as the ranking of the networks based on the intra-cluster densities of the blocking clusters of the networks.

We observe a moderately positive correlation (correlation coefficient values in the range of 0.5, ..., 0.7) between the CBI(BWC) values for the 40 real-world networks and the largest of the intra-cluster densities of the Louvain clusters in these networks. We also observe a moderately positive correlation between the CBI(DEG) and CBI(BWC) values for the 40 real-world networks, justifying our observations in Sect. 5.1 that the choice of the centrality metric could at most have a moderate impact on the CBI values of the networks. Putting together the observations in Sects. 5.1 and 5.2, we could conclude CBI(DEG) to be an ideal choice for quantifying as well as predicting/ranking the intra-cluster densities of the blocking clusters of complex real-world networks.

Distribution of the initial adopters in the Louvain clusters

For each real-world network and centrality metric (DEG and BWC), we determine the distribution of the initial adopter nodes corresponding to the f_{IA}^{\min} value (whose corresponding threshold fraction of adopted neighbors is $q_{step} = 1 - \text{CBI}$) in the clusters computed using the Louvain algorithm. Table 5 displays the number of initial adopters (corresponding to the f_{IA}^{\min} value) in each of the Louvain clusters for the real-world networks. We refer to "Fraction of IA clusters (f_{IA}^{Clusters})" as the ratio of the number of clusters that have at least one initial adopter node in them and the total number of clusters. We use the notations $f_{IA, \text{DEG}}^{\text{Clusters}}$ and $f_{IA, \text{BWC}}^{\text{Clusters}}$ to respectively indicate the fraction of IA clusters observed with respect to the DEG and BWC centrality metrics. Figure 9i plots the $f_{IA, \text{DEG}}^{\text{Clusters}}$ and $f_{IA, \text{BWC}}^{\text{Clusters}}$ in the decreasing order of their values for the real-world networks: the $f_{IA, \text{DEG}}^{\text{Clusters}}$ and $f_{IA, \text{BWC}}^{\text{Clusters}}$ values were equal to 1.00 (implying there was at least one initial adopter in each cluster) for respectively 19 and 27 of the 40 real-world networks, and

Table 5 Distribution of the initial adopter nodes (chosen based on the DEG and BWC centrality metrics) in the Louvain clusters of the real-world networks

#	Net. code	# Initial adopters/cluster deg-based selection	$f_{IA,DEG}^{Clusters}$	#	Net. code	# Initial adopters/cluster BWC-based selection	$f_{IA,BWC}^{Clusters}$
1	ADJ	3, 5, 3, 3, 2, 2	1.00	1	ADJ	5, 6, 5, 3, 3, 5	1.00
2	AKN	1, 2, 1, 1, 1, 0	0.83	2	AKN	1, 2, 1, 1, 1, 0	0.83
3	JBN	6, 15, 20, 4	1.00	3	JBN	21, 25, 29, 6	1.00
4	CBN	4, 10, 6, 3	1.00	4	CBN	5, 7, 5, 3	1.00
5	CEN	2, 6, 5, 10, 3, 18	1.00	5	CEN	2, 6, 7, 23, 12, 22	1.00
6	CMN	1, 0, 0, 0, 0, 0, 0, 0	0.11	6	CMN	10, 6, 12, 18, 2, 6, 10, 3, 8	1.00
7	CLN	10, 3, 3, 4, 2, 6, 3	1.00	7	CLN	3, 1, 3, 3, 3, 2, 2	1.00
8	CFN	1, 1, 1, 6, 2, 4, 12	1.00	8	CFN	0, 1, 0, 1, 0, 2, 2	0.57
9	CSA	2, 0, 0, 0	0.25	9	CSA	2, 0, 0, 0	0.25
10	DON	3, 0, 4, 3	0.75	10	DON	2, 2, 4, 5	1.00
11	DLN	0, 0, 0, 3, 0	0.20	11	DLN	1, 0, 2, 3, 0	0.60
12	EUT	2, 3, 2, 1, 2, 1, 1, 1	1.00	12	EUT	3, 1, 2, 1, 5, 1, 2, 1	1.00
13	FBN	0, 1, 0, 0, 1, 1, 2, 8	0.63	13	FBN	2, 6, 1, 1, 9, 4, 8, 15	1.00
14	FHT	0, 2, 2, 0	0.50	14	FHT	1, 4, 3, 5	1.00
15	BFW	1, 2, 0, 4, 0	0.60	15	BFW	13, 28, 13, 9, 7	1.00
16	FFW	8, 7, 7, 19	1.00	16	FFW	12, 15, 21, 24	1.00
17	FTC	1, 4, 2, 4	1.00	17	FTC	1, 3, 2, 3	1.00
18	GDN	1, 1, 1, 1, 1, 1, 1, 1	1.00	18	GDN	1, 1, 1, 1, 1, 1, 1, 1	1.00
19	ISP	6, 43, 36, 32	1.00	19	ISP	11, 11, 12, 3	1.00
20	JDN	2, 1, 3, 3, 3, 0, 0, 3, 0, 0, 7	0.64	20	JDN	38, 54, 73, 44, 29, 2, 8, 10, 26, 22, 2, 51, 1	1.00
21	KCN	0, 0, 0, 1	0.25	21	KCN	3, 4, 1, 4	1.00
22	KFP	0, 2, 0, 0, 0	0.20	22	KFP	0, 2, 0, 1, 1	0.60
23	LMN	0, 0, 0, 0, 1, 0	0.17	23	LMN	0, 0, 0, 0, 1, 0	0.17
24	MCE	2, 3, 15, 11, 10	1.00	24	MCE	2, 3, 12, 9, 9	1.00
25	MMN	1, 1, 1, 2, 2	1.00	25	MMN	1, 0, 2, 2, 1	0.80
26	PDL	1, 1, 1, 1, 2, 1, 0, 0, 0	0.67	26	PDL	1, 1, 1, 1, 2, 1, 0, 0, 0	0.67
27	PBN	1, 4, 0, 5	0.75	27	PBN	4, 7, 1, 6	1.00
28	PSN	22, 4, 15, 14, 1	1.00	28	PSN	24, 5, 12, 6, 7	1.00
29	PFN	4, 2, 2, 2, 1, 2, 2, 2	1.00	29	PFN	0, 1, 0, 1, 0, 0, 0, 0	0.25
30	RVU	6, 12, 7, 4	1.00	30	RVU	5, 10, 8, 3	1.00
31	SJN	2, 4, 2, 5, 3, 1, 2	1.00	31	SJN	1, 1, 1, 2, 0, 0, 1	0.71
32	SPR	8, 1, 4, 1, 3, 0	0.83	32	SPR	15, 2, 12, 4, 9, 6	1.00
33	SWC	2, 1, 2, 1, 0	0.80	33	SWC	4, 2, 5, 3, 3	1.00
34	TEN	1, 0, 0, 0, 0	0.20	34	TEN	0, 0, 1, 0, 0	0.20
35	UKF	2, 1, 3, 4	1.00	35	UKF	12, 1, 15, 13	1.00
36	APN	39, 18, 13, 0, 4, 3, 2	0.86	36	APN	3, 7, 2, 1, 2, 2, 2	1.00
37	USS	1, 0, 0, 0, 0, 1	0.33	37	USS	5, 5, 4, 4, 4, 5	1.00
38	WGA	0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1	0.77	38	WGA	0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1	0.54
39	RHF	16, 23, 18, 6, 18, 21	1.00	39	RHF	20, 27, 20, 8, 16, 21	1.00
40	YTN	2, 4, 1, 3, 0, 2, 0, 2, 1, 1, 6, 0, 1, 1, 2, 0, 0	0.70	40	YTN	2, 4, 2, 4, 0, 2, 0, 3, 1, 1, 5, 1, 1, 1	0.85

greater than 0.5 for respectively 31 and 36 of the real-world networks. From both Fig. 9-(i, ii), we could also infer that for a majority of the real-world networks (35 of the 40 real-world networks), the $f_{IA,BWC}^{Clusters}$ values are greater than or equal to the $f_{IA,DEG}^{Clusters}$ values.

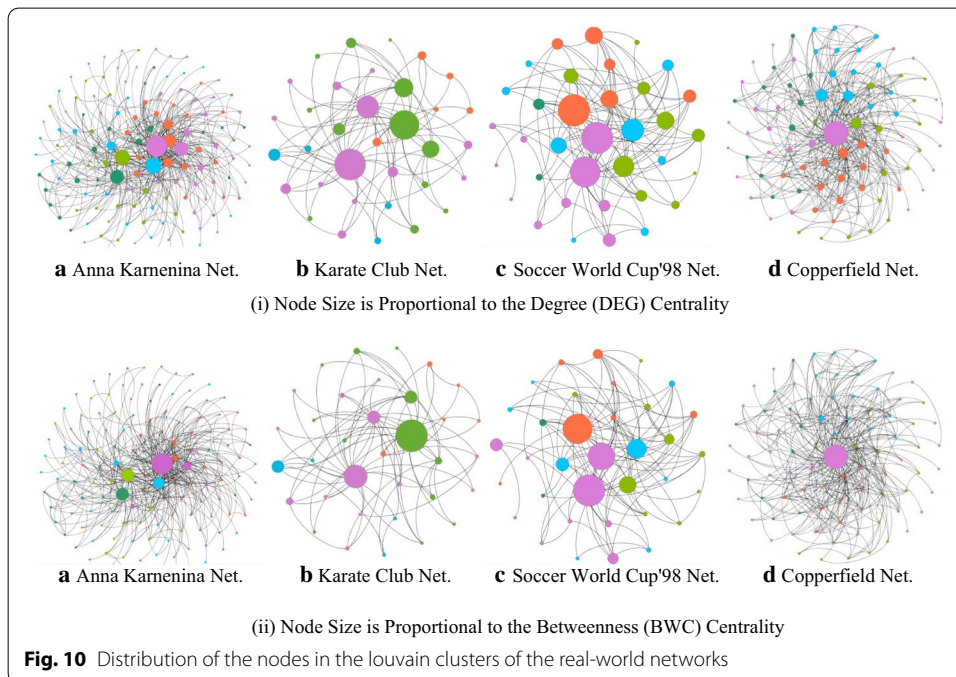
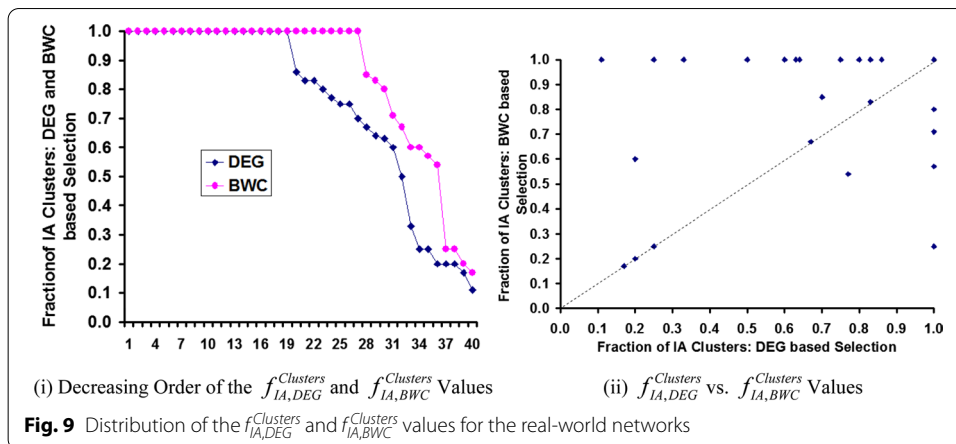
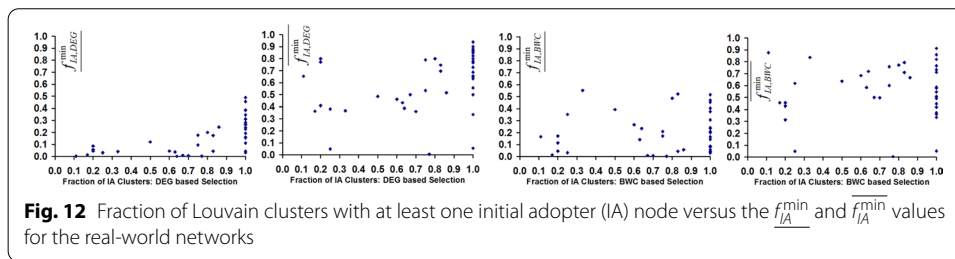
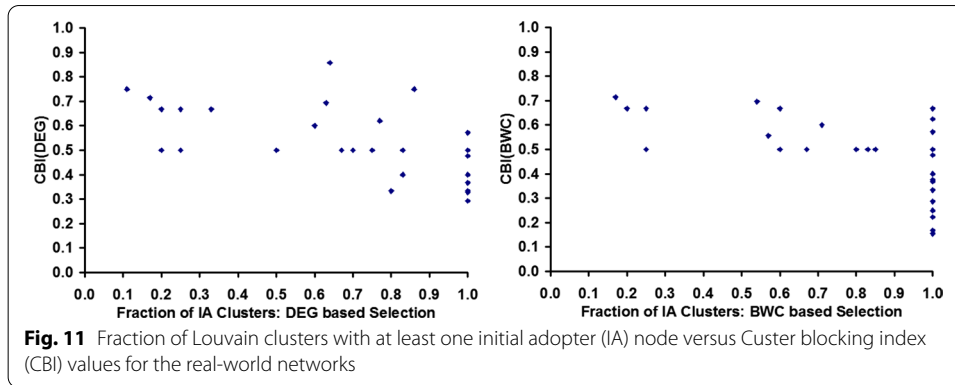


Figure 10 shows the distribution of the nodes (nodes with larger DEG and BWC values are bigger in size) in the different Louvain clusters of the four sample real-world networks that were shown in Fig. 3 (*Step Function Patterns Observed for the q versus Distribution for Real-World Networks with respect to Degree Centrality*). We observe nodes with larger DEG (in Fig. 10-i) and BWC (in Fig. 10-ii) values (that are candidates to be selected as the initial adopter nodes) to be distributed in all the Louvain clusters. Though we just present a sample of the real-world networks in Fig. 10-(i) and (ii), from Figs. 9 and 10, we could confidently conclude that the initial adopters chosen on the basis of the DEG or BWC metrics are not concentrated in just one cluster and are spread over multiple clusters of the real-world networks.

We now analyze the impact of the fraction of IA clusters on the CBI, $\overline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$ values observed for the real-world networks. Figure 11 presents plots of $\overline{f_{IA,DEG}^{Clusters}}$ versus



CBI(DEC) and $f_{IA,BWC}^{clusters}$ versus CBI(BWC) measures, wherein we observe only at most a moderate correlation (Pearson's correlation coefficient less than 0.60) between the measures. This implies that the CBI value for a real-world network is not heavily dependent on the presence or absence of one or more initial adopters in any particular cluster of the real-world networks. Likewise, Fig. 12 presents plots of $f_{IA,DEC}^{clusters}$ versus the f_{IA}^{min} values and the f_{IA}^{min} values for DEC and BWC-based initial adopter selection (denoted as $f_{IA,DEC}^{min}$, $f_{IA,BWC}^{min}$, $f_{IA,DEC}^{clusters}$ and $f_{IA,BWC}^{clusters}$ in Fig. 12), wherein we only observe weak-moderate correlation (Pearson's correlation coefficients in the range of 0.4 to 0.6) in the case of DEC and no correlation (Pearson's correlation coefficients in the range of 0.0 to 0.1) in the case of BWC.

Analysis of the computation times of the algorithms for the real-world networks

In this sub section, we discuss the computation times observed for the binary search approach of Algorithm 3 to determine the q_{step} value for a real-world network with respect to a centrality metric (used to choose the initial adopters) as well as for the binary search approach of Algorithm 2 (which uses Algorithm 1 for running iterative information cascade) to determine the f_{IA}^{min} and $f_{IA}^{clusters}$ values needed to operate the network to accomplish complete information cascade at q_{step} and the next measurable value of q_{step} respectively. These computation times are reported in Table 6. All the three algorithms (Algorithms 1, 2 and 3) are implemented in Java and run on a desktop Windows 7 computer (Intel i7-2620 M CPU @ 2.70 GHz with 8 GB RAM). The computation times

Table 6 Computation times for the q_{step} , f_{IA}^{min} and $\overline{f_{IA}^{min}}$ values for the real-world networks

Computation times (milliseconds) with respect to DEG centrality					Computation times (milliseconds) with respect to BWC centrality				
#	Net. code	q_{step}	f_{IA}^{min}	$\overline{f_{IA}^{min}}$	#	Net. code	q_{step}	f_{IA}^{min}	$\overline{f_{IA}^{min}}$
1	ADJ	22.2	14.4	3.6	1	ADJ	34.9	16.5	4.5
2	AKN	15.8	4.5	2.2	2	AKN	15.6	5.1	2.4
3	JBN	42.8	20.3	3.6	3	JBN	70.1	20.9	4.5
4	CBN	12.6	2.5	0.7	4	CBN	13.8	2.6	2.0
5	CEN	161.9	64.1	23.0	5	CEN	175.3	45.9	11.1
6	CMN	84.2	25.1	9.4	6	CMN	62.5	12.7	6.4
7	CLN	9.5	1.6	0.5	7	CLN	9.9	2.9	1.1
8	CFN	2.8	0.8	0.4	8	CFN	5.9	1.4	0.3
9	CSA	2.4	0.3	0.3	9	CSA	1.8	0.3	0.2
10	DON	2.7	1.0	0.3	10	DON	1.6	0.6	0.2
11	DLN	1.0	0.3	0.1	11	DLN	1.1	0.2	0.2
12	EUT	81.3	10.4	10.5	12	EUT	78.0	18.7	9.4
13	FBN	167.5	35.7	14.0	13	FBN	84.3	21.3	5.7
14	FHT	1.1	0.2	0.1	14	FHT	0.6	0.3	0.2
15	BFW	19.2	8.3	1.8	15	BFW	14.4	2.9	1.2
16	FFW	29.6	7.9	1.8	16	FFW	20.3	4.1	1.8
17	FTC	2.1	0.4	0.2	17	FTC	3.7	0.5	0.2
18	GDN	8.4	2.6	1.2	18	GDN	7.5	2.1	0.8
19	ISP	77.9	18.2	5.5	19	ISP	177.5	32.1	6.9
20	JDN	1032.1	146.6	84.2	20	JDN	536.7	136.6	58.0
21	KCN	1.1	0.2	0.1	21	KCN	0.8	0.1	0.1
22	KFP	1.6	0.3	0.1	22	KFP	1.6	0.2	0.2
23	LMN	7.9	1.1	0.7	23	LMN	6.9	0.9	0.3
24	MCE	11.3	2.0	1.1	24	MCE	13.0	1.8	1.0
25	MMN	1.2	0.2	0.1	25	MMN	0.7	0.2	0.1
26	PDL	152.4	66.4	14.3	26	PDL	148.6	43.4	13.8
27	PBN	8.1	3.0	0.7	27	PBN	6.7	2.7	0.6
28	PSN	172.2	48.1	4.5	28	PSN	214.0	30.9	7.8
29	PFN	2.0	0.6	0.2	29	PFN	4.4	0.7	0.4
30	RVU	467.0	132.0	50.7	30	RVU	484.9	209.1	37.3
31	SJN	2.8	0.6	0.3	31	SJN	6.5	1.3	0.7
32	SPR	8.3	2.3	0.8	32	SPR	3.4	0.6	0.4
33	SWC	1.2	0.2	0.1	33	SWC	0.7	0.3	0.1
34	TEN	0.5	0.1	0.1	34	TEN	0.5	0.1	0.1
35	UKF	10.8	1.9	0.6	35	UKF	4.3	1.0	0.4
36	APN	38.8	7.5	4.9	36	APN	70.9	11.2	3.5
37	USS	3.2	0.6	0.3	37	USS	1.7	0.3	0.2
38	WGA	2109.5	283.9	249.2	38	WGA	1903.3	295.1	292.3
39	RHF	77.4	14.6	6.7	39	RHF	37.5	6.7	3.1
40	YTN	7047.3	3556.1	612.1	40	YTN	6616.5	3227.8	480.9

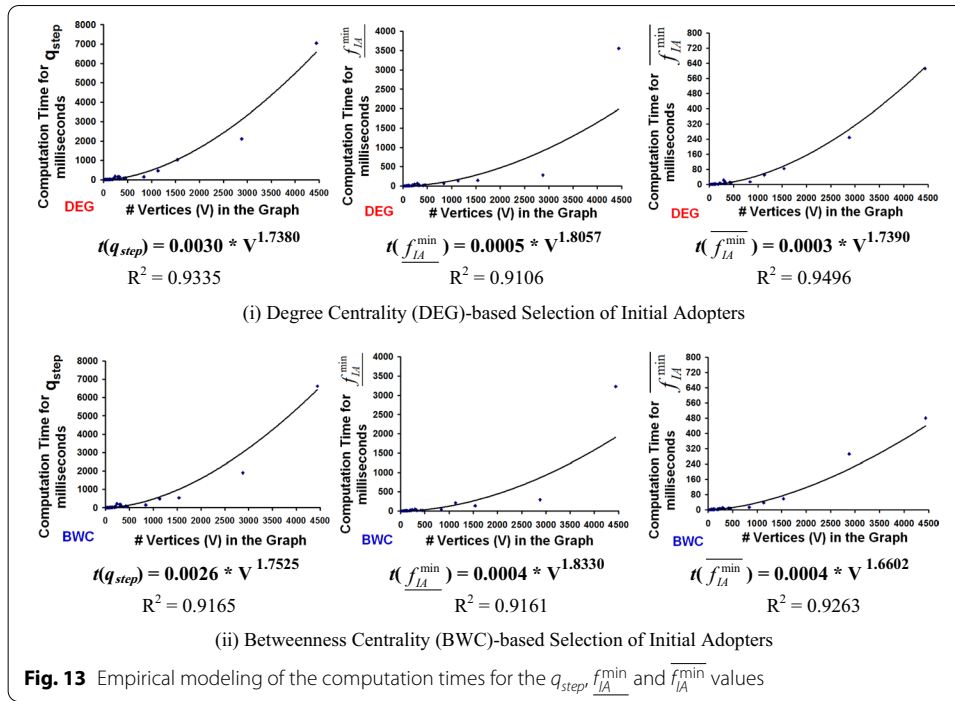
of the centrality metrics (DEG and BWC) of the vertices in the real-world networks are not accounted for in the computation times reported in Table 6.

For both the centrality metrics and for each of the 40 real-world networks, we observe the computation times for $\underline{f_{IA}^{min}}$ to be greater than that of $\overline{f_{IA}^{min}}$. Though the

number of iterations for Algorithm 2 (*Binary Search Algorithm to Determine the Minimum Number of Initial Adopters for a Given Threshold Fraction of Adopted Neighbors: q*) to compute $\underline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$ is $\ln(V)$, where V is the number of vertices in the network, the relatively larger computation time for $\underline{f_{IA}^{\min}}$ could be attributed to the larger number of iterations of Algorithm 1 (*Iterative Algorithm to Conduct Information Cascade for a Given Set of Initial Adopters: IA and a Threshold Fraction of Adopted Neighbors: q*) for each iteration of Algorithm 2. Remember that in Algorithm 1, we conclude that complete information cascade is not accomplishable for a particular value of q and IA if there are no newly adopted vertices in the latest iteration. When operated at q (threshold fraction of adopted neighbors) values less than or equal to q_{step} and a smaller sized set of initial adopters, it is more likely that the number of newly adopted vertices in each iteration of Algorithm 1 would be greater than 0: leading to relatively more iterations before we could conclude whether complete information cascade is accomplishable or not for a particular value of q . On the other hand, when operated at q values above q_{step} and a larger sized set of initial adopters, we could decide whether complete information cascade is accomplishable or not for a particular value of q by going through relatively fewer iterations of Algorithm 1. When operated with a search space of 0.05 and termination threshold $\varepsilon = 0.001$, Algorithm 3 (*Binary Search Algorithm to Analyze the Jump Zone of a q versus Distribution*) runs Algorithm 2 exactly six times (see Sect. 4.1 for more details). Of course, as discussed above, the computation time for Algorithm 2 depends on the q value considered. Hence, we expect the computation time of Algorithm 3 for any real-world network and centrality metric to be greater than the computation time reported for $\underline{f_{IA}^{\min}}$ (corresponding to one run of Algorithm 2) for the particular real-world network but less than or equal to six times the computation time for $\underline{f_{IA}^{\min}}$.

An interesting observation in Table 6 is that for more than half of the real-world networks, the computation times reported for any of the three measures (q_{step} , $\underline{f_{IA}^{\min}}$, $\overline{f_{IA}^{\min}}$) are relatively lower when the BWC metric is used to choose the initial adopters compared to the DEG metric. This is because the q_{step} values for at least 50% of the real-world networks with respect to the BWC metric are observed to be larger than those observed with respect to the DEG metric. Based on the earlier discussion in this subsection, Algorithm 1 takes relatively less time to decide whether or not complete information cascade is accomplishable when operated at a larger q value. It is thus logical to observe relatively lower computation times for the q_{step} , $\underline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$ values for the real-world networks when the initial adopters are chosen with respect to the BWC metric. Note that, as mentioned earlier, we did not take into consideration the computation times for the centrality metrics (DEG or BWC) that are used to choose the set of initial adopters. It is well-known in the literature that BWC is a computationally-heavy metric (Meghanathan 2017b) and DEG is a computationally-light metric. If we were to incorporate the computation times of the centrality metrics in the computation times of the q_{step} , $\underline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$ measures, it would skew the values and we will not be able to make any inferential observation.

We seek to model the computation times for each of the q_{step} , $\underline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$ measures reported in Table 6 as a function of the number of nodes (V) in the network. After trying different models, we observe the model for computation time $t = a^*V^b$ to be the most



closest fit with R^2 values above 0.90, wherein the coefficients a and b vary depending on the measure (q_{step} , f_{LA}^{\min} and f_{IA}^{\min}) and centrality metric (DEG or BWC) considered. Figure 13-(i) and (ii) present plots of the # vertices (V) versus the computation time (t) values for each of the above three measures with respect to DEG and BWC respectively; the models that we fit for each of these six plots are shown below the plots. The models for the computation times (polynomial functions of the number of vertices in the graph, with degree less than 2) clearly indicate the scalability of the binary search-based algorithms 2 and 3 proposed in this research.

The theoretical time complexities of the well-known clustering algorithms in the literature range from $O(E)$: Louvain algorithm (Blondel et al. 2008) to $O(E^2V)$: Girvan-Newman algorithm (Newman 2004), where V and E are respectively the number of vertices and edges in a network. For dense graphs (Cormen et al. 2009): $E = O(V^2)$ and for sparse graphs (Cormen et al. 2009): $E = O(V)$. The theoretical time complexity of the proposed binary search algorithms 2 and 3 (see Sects. 22–4 is $O(E \cdot V \log V)$, which falls in between the extremes of $O(E)$ and $O(E^2V)$. Moreover, the actual time complexities (empirical models) of algorithms 2 and 3 (polynomials with degree greater than one, but less than two, as indicated in Fig. 13) are all observed to be much lower than $E \cdot V \log V$, the theoretical upper bound. Thus, the empirical models of the computation times for the q_{step} , f_{LA}^{\min} and f_{IA}^{\min} measures presented in this sub section clearly demonstrate the suitability of the proposed binary search algorithms to run for very larger networks (with the number of nodes much greater than the ones studied in this paper) and justify our claim that the CBI values be considered a (perhaps computationally-light) quantitative measure of the intra-cluster density of the blocking clusters of a network. Even for a network of 100,000 nodes, the computation times (on a

regular desktop computer with Intel i7-2620 M CPU @ 2.70 GHz with 8 GB RAM) for the q_{step} , f_{IA}^{\min} and $\overline{f_{IA}^{\min}}$ values would be close to 25 min, 9 min and 3 min respectively for DEG-based selection of initial adopters and would be close to 25 min, 10 min and 2 min respectively for BWC-based selection of initial adopters.

Related work

In Jalili and Perc (2017), the authors quantified the spreading influence of a node as the fraction of nodes that adopt a decision based on the decision adopted by the node. They observed a positive correlation between the degree (DEG) and betweenness (BWC) centrality metrics versus the spreading influence of a node. In Yang and Leskovec (2010), the authors observed that for effective dissemination of information, the connections of the initial adopter nodes are more important than the number of initial adopters. In Ghasemiesfeh et al. (2013), the authors observed that for the complex contagion model that is also used in this paper (a node adopts a decision only when the fraction of neighbors that have adopted the decision is greater than or equal to a threshold), the diffusion speed depends on the distribution of the weak ties [nodes with larger BWC are considered to be weak ties in a network (Easley and Kleinberg 2010)]. In Bakshy et al. (2012), the effect of tie strength on information sharing and diffusion in Facebook was studied. It was observed that strong ties (friends who interacted more frequently) were effective in information cascade within clusters, whereas weak ties (friends who did not interact more frequently) were effective in information cascade across clusters. It was also observed that the probability of sharing of a link by a user increases with the number of sharing friends. In Watts (2002), it was observed that increased heterogeneity in the ' q ' values (threshold fraction of adopted neighbors needed for a node to adopt a decision) reduces the chances of success in accomplishing complete information cascade (note that we use the same value of ' q ' for all the nodes in the network), whereas increased heterogeneity among nodes with respect to degree makes a network relatively less vulnerable to information cascade. In a prior work (Meghanathan 2019), we observed that real-world networks are more dissortative with respect to the remaining degree of the vertices (one less than the degree of a vertex) and the betweenness centrality metric, and more assortative with respect to the eigenvector (Bonacich 1987) and closeness (Freeman 1979) centrality metrics. As a result, it is very less likely that the strategy of selecting nodes in the decreasing order of their DEG or BWC values to form the set of initial adopters would lead to all the nodes from the same cluster being selected. Using all of the above observations as guidelines, we decided to use the DEG and BWC metrics as the basis to choose the initial adopters.

Though the terminologies "information diffusion" and "information cascade" are sometimes interchangeably used in the literature, a recent work (Buskens 2019) clearly distinguished information diffusion from information cascade [referred to as information appreciation in Buskens (2019)] as well as explained the issue of selection of initial adopters for these two phenomena. *Information diffusion* is a phenomenon in which the information has to just reach all the nodes in the network; whereas *Information cascade* is a phenomenon wherein the nodes are required to adopt/accept the information and propagate further. The complex contagion model used in this paper is one of the commonly used models for information cascade. For information cascade, there needs to be

some redundancy in the channels through which information propagates (to ensure a threshold fraction of the neighbors of a node have adopted/accepted the information before the node does so); whereas for information diffusion, redundancy in the number of channels through which information propagates could actually slow down the spread. For faster information diffusion with minimal redundancy, it would be more prudent to choose the initial adopters spread over in different clusters. On the other hand, for complete information cascade (which is the focus of this paper), due to the need for redundancy to accept the information and propagate further, it would be more effective to include one or more bridge nodes of the clusters to be part of the set of initial adopters.

In Chesney (2017), the author came up with the notion of the "cascade capacity of a node" which in the context of this paper is the maximum q value that we could employ for a network and accomplish complete information cascade by starting with the node as the only initial adopter. Like centrality metrics, each node would have a different cascade capacity depending on its position in the network vis-a-vis its neighbors. A simulation-based iterative procedure was proposed in Chesney (2017) to determine the cascade capacities of the individual nodes in a network. The cascade capacity of a node was observed to be weakly correlated with centrality metrics such as DEG and BWC that are typically used to choose the initial adopters. Like the DEG and BWC metrics, the cascade capacity of a node is a valuable information that could also be considered to choose the set of initial adopters and we intend to explore this in our future research. However, the time complexity to determine the cascade capacity of a node is not formally evaluated in Chesney (2017). As the time complexity will be overwhelming if one were to adopt a brute force approach to determine the cascade capacity of the nodes on a node-by-node basis, as part of future work, we plan to investigate the relationship between the cascade capacity of the nodes and the CBI value for a network and accordingly develop binary search approaches exploiting any such relationship that might exist between the two measures and determine the cascade capacity of the nodes in an efficient manner.

In Watts and Dodds (2007), it was observed that high-degree nodes play a critical role to choose an appropriate value for the fraction of initial adopters needed to induce complete information cascade in random networks. Though the above work was done primarily for random networks, the theoretical observation made in Watts and Dodds (2007) is also observed to hold good for the real-world networks analyzed in our research. The high-degree bridge nodes of the clusters are the ones that appear to be the stumbling block for information cascade to penetrate through the clusters and we need to ramp up the fraction of initial adopters when the threshold fraction (q) of adopted neighbors exceeds q_{step} . As a result, inclusion of high-degree nodes in the set of initial adopters turns out to be a relatively more effective strategy (compared to the inclusion of nodes with high BWC) to quantify the intra-cluster density of the blocking clusters of the real-world networks. In another related work (Watts and Dodds 2007) on random networks, the author observed that heterogeneous thresholds (different q values for the nodes) are more likely to result in complete information cascade. As part of future work, we plan to evaluate the impact of heterogeneous threshold values for the fraction of adopted neighbors on the CBI values of real-world networks.

Information cascade has been observed to play a significant role in sequential voting mechanisms such as presidential primaries and roll-call voting (Knight and Schiff

2010), wherein information about the previously cast votes are revealed to the voter and accordingly the voter finalizes his/her choice of the candidate that could be even different from their personal preference/initial choice. In a recent work (Tump et al. 2020), the authors modeled the dynamics of social decision-making process leading to information cascade and observed the drift rate from the personal choice to the majority choice exhibited a convex increase for smaller majority size and a concave increase for medium and larger majority sizes. The impact of the decision made by the initial adopters on the final outcome of a sequential unanimous decision-making information cascade phenomenon was also experimentally studied in Anderson and Holt (1997) involving several test subjects. Each test subject was asked to predict the predominant color of a collection of balls in an urn after looking at the color of a randomly drawn ball by the test subject as well as considering the colors of the balls drawn by the earlier test subjects. It was observed that if the initial few decisions coincide, the subsequent test subjects also took the same decision as the earlier test subjects, irrespective of the color of the ball drawn by the test subjects. Similar experiments were also designed in Alevy et al. (2007) and Mori et al. (2013). Such experiments motivated us to develop an algorithm to determine the minimum fraction of nodes to be chosen as initial adopters to accomplish complete information cascade (i.e., adopt a unanimous decision) for a given threshold fraction (q) of adopted neighbors.

In a recent work (Hisakado and Mori 2009), voter model dynamics was studied in the context of information cascade: the authors observed that in a population mix of independent voters (who vote on their own) and copycat voters (who vote probabilistically based on the number of votes polled so far for each candidate), the distribution of the voting rate is observed to go through a complete phase transition (i.e., from a binomial distribution to a beta distribution) only when the number of votes seen by the copycat voters is significantly high (theoretically, infinity). On the other hand, if the fraction of copycat voters is $1/2$ or more, the voting rate converges more slowly. The voting style of the copycat voters in the above work is similar to the adoption behavior of the nodes under the complex contagion model for information cascade. In Watts and Dodds (2007), the distribution of the cascade size (the number of nodes that adopt the decision of the initial adopters) versus the fraction of nodes to be chosen as initial adopters was observed to be bimodal for synthetic networks generated per the Watts model (Watts 2002), indicating that there is a threshold value for the fraction of adopted neighbors that connects the two distributions. The sudden significant increase observed in our research in the fraction of initial adopters needed for complete information cascade with increase in q beyond q_{step} also resembles the phase transition and bimodal distribution reported in the above theoretical works.

Conclusions and future work

The high-level contributions of this paper are the following: We proposed a binary search algorithm to determine the minimum fraction of nodes to be used as initial adopters (f_{IA}^{\min}) to accomplish complete information cascade for a given threshold fraction (q) of adopted neighbors. We analyzed a suite of 40 real-world networks of diverse domains and observed the q versus f_{IA}^{\min} distribution to exhibit a step function pattern for 37 of these networks and identified a jump zone within which the f_{IA}^{\min} value

spiked from $\underline{f_{IA}^{\min}}$ to $\overline{f_{IA}^{\min}}$ (with a median difference of 0.44 and 0.32 respectively when degree and betweenness centrality metrics are used to choose the initial adopters). We proposed a second binary search algorithm (that makes use of the first binary search algorithm) to determine the ' q ' value (referred to as q_{step}) beyond which we needed to increase the minimum fraction of nodes as initial adopters from $\underline{f_{IA}^{\min}}$ to $\overline{f_{IA}^{\min}}$ in order to accomplish complete information cascade. We propose that the " $1 - q_{step}$ " value (referred to as the Cluster Blocking Index: CBI) for a network be considered as a measure of the intra-cluster density of the blocking cluster of the network as any further increase in q (beyond q_{step}) would make the intra-cluster density of at least one cluster (the blocking cluster) to become greater than $1 - q$. We will not be able to penetrate such a cluster with $\underline{f_{IA}^{\min}}$ fraction of initial adopters and will need to increase the minimum fraction of initial adopters to $\overline{f_{IA}^{\min}}$ by including one or more nodes of the blocking cluster to be part of the set of initial adopters. The CBI value for a network could be used to decide the appropriate value for the threshold fraction of adopted neighbors needed to facilitate information cascade or stop infection spread. For networks with larger CBI values: complete information cascade can be accomplished only when operated with lower q values (i.e., nodes are forced to take a decision when only fewer of their neighbors have made the decision); an infection spread can be avoided by operating with larger q values (equivalent to vaccinating the bridge nodes). We observe a majority of the 40 real-world networks to incur larger CBI values (of 0.50 or more) with respect to both the degree (DEG) and betweenness (BWC) metrics, the centrality metrics used to choose the initial adopter nodes. We observe the fraction of IA clusters (fraction of Louvain clusters with one or more initial adopter nodes) is 1.00 for several real-world networks and has only at most a moderate correlation with the CBI values determined for the networks.

The CBI value for a network is not dependent on any particular clustering algorithm as well as there is no need to run any clustering algorithm to determine the clusters, evaluate their intra-cluster densities to identify the blocking cluster and thereby decide on the CBI value for a network. Note that the two-phase approach explained in Sects. 3 and 4 to determine the CBI value for a network does not use any clustering algorithm. We expect the CBI value to serve as the upper bound for the intra-cluster densities of the clusters determined by any clustering algorithm. We verified our claim in this paper by determining the intra-cluster densities of the clusters in 40 real-world networks using the well-known Louvain community detection algorithm. While the CBI(DEG) values were observed to serve as upper bound for the intra-cluster densities of the Louvain clusters for 32 of the 40 real-world networks, the CBI(BWC) values were observed to serve as upper bound for 27 of the 40 real-world networks. The DEG metric has been observed to be effective in determining CBI values that could serve as upper bound for the intra-cluster densities of the clusters in the real-world networks as well as be used to predict the intra-cluster density of the blocking cluster of a network. We also validated the scalability of the proposed binary search algorithms for large network graphs by deriving empirical models for the actual time complexities (polynomial functions of degree greater than one, but less than two) encountered to determine the three measures: q_{step} , $\underline{f_{IA}^{\min}}$ and $\overline{f_{IA}^{\min}}$.

We hypothesize that smaller clusters in a network are more likely to become the blocking clusters for complete information cascade as the fraction of alien neighbors

(fraction of neighbors that are outside the cluster) for the bridge nodes in smaller clusters are likely to be lower than the fraction of alien neighbors for the bridge nodes in larger clusters. As part of future work, we plan to investigate the role of cluster size in the intra-cluster densities and the CBI values of the network. Bridge nodes are the entry points to penetrate through a cluster. Though there exists centrality metrics to quantify the "bridgeness" of the nodes in networks (Jensen et al. 2015), these centrality metrics typically identify nodes that connect two or more clusters, but are not part of either of the clusters. As part of future work, we also plan to work on developing a centrality metric that quantifies the extent to which a node can let information cascade to successfully penetrate through a cluster. In addition to the above, for future work, we also plan to explore the open research problems mentioned in the Related Work section.

Abbreviations

f_{IA}^{\min} : Minimum fraction of nodes in a network to be used as initial adopters; q : Threshold fraction of adopted neighbors; q_{step} : The threshold fraction of adopted neighbors beyond which the f_{IA}^{\min} value exhibits a sharp increase; $\overline{f_{IA}^{\min}}$: The f_{IA}^{\min} value corresponding to q_{step} ; f_{IA}^{\min} : The f_{IA}^{\min} value corresponding to the next measurable value of q beyond q_{step} ; DEG: Degree centrality; BWC: Betweenness centrality; IA_{min} : The minimum number of nodes to be used as initial adopters for a given q ; CBI: Cluster Blocking Index; CBI(DEG): Cluster Blocking Index of a network with the degree centrality metric used to choose the initial adopters; CBI(BWC): Cluster Blocking Index of a network with the betweenness centrality metric used to choose the initial adopters; IA: Set of initial adopters; LI: Left index of the binary search algorithm; MI: Middle index of the binary search algorithm; RI: Right index of the binary search algorithm.

Acknowledgements

Not applicable

Authors' contributions

Natarajan Meghanathan was solely responsible for all work presented in this paper. All authors read and approved the final manuscript.

Funding

This work was partially supported through the sub contract received from University of Virginia titled: Global Pervasive Computational Epidemiology, with the National Science Foundation as the primary funding agency. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the funding agencies.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The author declares that they have no competing interests.

Received: 29 August 2020 Accepted: 24 November 2020

Published online: 04 December 2020

References

- Alevy JE, Haigh MS, List JA (2007) Information cascades: evidence from a field experiment with financial market professionals. *J Finance* 62(1):151–180
- Anderson LR, Holt CA (1997) Information cascades in the laboratory. *Am Econ Rev* 87(5):847–862
- Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: Proceedings of the 21st international conference on world wide web, Lyon, France, pp 519–528
- Batagelj V, Mrvar A (2006) *Pajek datasets*. <http://vlado.fmf.uni-lj.si/pub/networks/data/>
- Bhardwaj N, Yan KK, Gerstein MB (2010) Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. *Proc Natl Acad Sci USA* 107(15):6841–6846
- Blagus N, Subelj L, Bajec M (2012) Self-similar scaling of density in complex real-world networks. *Phys A* 391(8):2794–2802
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):1–12
- Bonacich P (1987) Power and centrality: a family of measures. *Am J Sociol* 92(5):1170–1182
- Buskens V (2019) Spreading information and developing trust in social networks to accelerate diffusion of innovations. In: Proceedings of the 33rd European federation of food science and technology international conference, Rotterdam, Netherlands

- Cadrillo A, Gomez-Gardenes J, Zanin M, Romance M, Papo D, Pozo F, Boccaletti S (2013) Emergence of network features from multiplexity. *Sci Rep* 3:1344
- Chesney T (2017) The cascade capacity predicts individuals to seed for diffusion through social networks. *Syst Res Behav Sci* 34(1):51–61
- Cormen TH, Leiserson CE, Rivest RL, Stein C (2009) Introduction to algorithms, 3rd edn. MIT Press, Cambridge
- Cross RL, Parker A, Cross R (2004) *The hidden power of social networks: understanding how work really gets done in organizations*, 1st edn. Harvard Business Review Press, London
- de Nooy W (1999) A literary playground: literary criticism and balance theory. *Poetics* 26(5–6):385–404
- Duch J, Arenas A (2005) Communication detection in complex networks using extremal optimization. *Phys Rev E* 72:027104
- Easley D, Kleinberg J (2010) Networks, crowds, and markets: reasoning about a highly connected world, 1st edn. Cambridge University Press, Cambridge
- Freeman L (1977) A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41
- Freeman L (1979) Centrality in social networks: conceptual classification. *Soc Netw* 1(3):215–239
- Freeman LC, Webster CM, Kirke DM (1998) Exploring social structure using dynamic three-dimensional color images. *Soc Netw* 20(2):109–118
- Geiser P, Danon L (2003) Community structure in Jazz. *Adv Complex Syst* 6(4):563–573
- Gemmetto V, Barrat A, Cattuto C (2014) Mitigation of infectious disease at school: targeted class closure vs. school closure. *BMC Infect Dis* 14(695):1–10
- Gephi (2020) <https://gephi.org/>. Last accessed: 1, 2020
- Ghasemiasfeh G, Ebrahimi R, Gao J (2013) Complex contagion and the weakness of long ties in social networks: revisited. In: Proceedings of the 14th ACM conference on electronic commerce, Philadelphia, PA, USA, pp 507–524
- Grimmer J (2010) A Bayesian hierarchical topic mode for political texts: measuring expressed agendas in senate press releases. *Polit Anal* 18(1):1–35
- Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A (2003) Self-similar community structure in a network of human interactions. *Phys Rev E* 68:065103
- Heymann S (2009) CPAN-explorer, an interactive exploration of the Perl Ecosystem. Gephi Blog
- Hisakado M, Mori S (2009) Phase transition and information cascade in a voting model. *J Phys A Math Theor* 43(31):1–13
- Hummon NP, Doreian P, Freeman LC (1990) Analyzing the structure of the centrality-productivity literature created between 1948 and 1979. *Sci Commun* 11(4):459–480. <https://doi.org/10.1177/107554709001100405>
- Ikeda Y, Hasegawa T, Nemoto K (2010) Cascade dynamics on clustered network. *J Phys Conf Ser* 221(012005):1–6
- Isella L, Stehle J, Barrat A, Cattuto C, Pinton JF, Van den Broeck W (2011) What's in a crowd? Analysis of face-to-face behavioral networks. *J Theor Biol* 271(1):166–180
- Jalili M, Perc M (2017) Information cascades in complex networks. *J Complex Netw* 5(5):665–693
- Jensen P, Morini M, Karsai M, Venturini T, Vespignani A, Jacomy M, Cointet JP, Merckle P, Fleury E (2015) Detecting global bridges in networks. *J Complex Netw* 4(3):319–329
- Knight B, Schiff N (2010) Momentum and social learning in presidential primaries. *J Polit Econ* 118(6):1110–1150
- Knuth DE (1993) *The Stanford graphbase: a platform for combinatorial computing*, 1st edn. Addison-Wesley, Reading
- Krackhardt D (1999) The ties that torture: simmelian tie analysis in organizations. *Res Sociol Organ* 16:183–210
- Krebs V (2003) Proxy networks: analyzing one network to reveal another. *Bull Methodol Sociol* 79:61–40
- Loomis CP, Morales JO, Clifford RA, Leonard OE (1953) Turrialba social systems and the introduction of change. The Free Press, Glencoe, pp 45–78
- Lusseau D, Schneider K, Boisseau OJ, Haase P, Slooten E, Dawson SM (2003) The Bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behav Ecol Sociobiol* 54(3):396–405
- MacRae D (1960) Direct factor analysis of sociometric data. *Sociometry* 23(4):360–371
- Magnani M, Micenkova B, Rossi L (2013) Combinatorial analysis of multiple networks. [arXiv:1303.4986](https://arxiv.org/abs/1303.4986) [cs.SI]
- Marengo JL (1960) The sociometry reader. The Free Press, Glencoe, pp 534–547
- Meghanathan N (2017a) Complex network analysis of the contiguous United States graph. *Comput Inf Sci* 10(1):54–76
- Meghanathan N (2017b) A computationally-lightweight and localized centrality metric in lieu of betweenness centrality for complex network analysis. *Viet J Comput Sci* 4(1):23–38
- Meghanathan N (2019) Centrality and partial correlation coefficient-based assortativity analysis of real-world networks. *Comput J* 62(9):1247–1264
- Mori S, Hisakado M, Takahashi T (2013) Collective adoption of max–min strategy in an information cascade voting experiment. *J Phys Soc Jpn* 82(8):1–10
- Nepusz T, Petroczi A, Negyessy L, Bazso F (2008) Fuzzy communities and the concept of bridgeness in complex networks. *Phys Rev E* 77(1):016107
- Newman MEJ (2004) Detecting community structure in networks. *Eur Phys J B* 38:321–330
- Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E* 74(3):036104
- Newman MEJ (2010) *Networks: an introduction*, vol 1. Oxford University Press, Oxford
- Prokhorenkova L, Tikhonov A, Litvak N (2019) Learning clusters through information diffusion. In: Proceedings of the world wide web conference, San Francisco, CA, USA, pp 3151–3157
- Ramezani M, Khodadadi A, Rabiee HR (2018) Community detection using diffusion information. *ACM Trans Knowl Discov Data* 12(2):1–22
- Rodriguez MG, Leskovec J, Balduzzi D, Scholkopf B (2014) Uncovering the structure and temporal dynamics of information propagation. *Netw Sci* 2(1):26–65
- Rogers EM, Kincaid DL (1980) Communication networks: toward a new paradigm for research. Free Press, London
- Scannell JW, Blakemore C, Young MP (1995) Analysis of connectivity in the cat cerebral cortex. *J Neurosci* 15(2):1463–1483
- Schwimmer E (1973) Exchange in the social structure of the Orokaiva: traditional and emergent ideologies in the Northern District of Papua, C Hurst and Co-Publishers Ltd

- Subelj L, Bajec M (2012) Ubiquitousness of link-density and link-pattern communities in real-world networks. *Eur Phys J B* 85(1):1–11
- Tump AN, Pleskac TJ, Kurvers RHJM (2020) Wise or mad crowds? The cognitive mechanisms underlying information cascades. *Sci Adv* 6(29):0266
- Ulanowicz R, Donald D (2005) Network analysis of trophic dynamics in South Florida Ecosystems. US Geological Survey Program on the South Florida Ecosystem, pp 114–115
- Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci USA* 99(9):5766–5771
- Watts DJ, Dodds P (2007) Influentials, networks and public opinion information. *J Consum Res* 34(4):441–458
- White JG, Southgate E, Thomson JN, Brenner S (1986) The structure of the nervous system of the nematode *Caenorhabditis Elegans*. *Philos Trans B* 314(1165):1–340
- Yang J, Leskovec J (2011) Modeling information diffusion in implicit networks. In: Proceedings of the 2010 IEEE international conference on data mining, Sydney, Australia, pp 599–608
- Zachary WW (1977) An information flow model for conflict and fission in small groups. *J Anthropol Res* 33(4):452–473
- Zander CD et al (2011) Food web including metazoan parasites for a brackish shallow water ecosystem in Germany and Denmark. *Ecology* 92(10):2007

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
