

RESEARCH ARTICLE

# Learning from Bees: An Approach for Influence Maximization on Viral Campaigns

C. Prem Sankar<sup>1</sup>, Asharaf S.<sup>2</sup>, K. Satheesh Kumar<sup>1\*</sup>

**1** Department of Futures Studies, University of Kerala, Kariavattom, Kerala, India - 695 581, **2** IIITM-K, Thiruvananthapuram, Kerala, India - 695 581

\* [kskumar@keralauniversity.ac.in](mailto:kskumar@keralauniversity.ac.in)

## Abstract

Maximisation of influence propagation is a key ingredient to any viral marketing or socio-political campaigns. However, it is an NP-hard problem, and various approximate algorithms have been suggested to address the issue, though not largely successful. In this paper, we propose a bio-inspired approach to select the initial set of nodes which is significant in rapid convergence towards a sub-optimal solution in minimal runtime. The performance of the algorithm is evaluated using the re-tweet network of the hashtag #KissOfLove on Twitter associated with the non-violent protest against the moral policing spread to many parts of India. Comparison with existing centrality based node ranking process the proposed method significant improvement on influence propagation. The proposed algorithm is one of the hardly few bio-inspired algorithms in network theory. We also report the results of the exploratory analysis of the network kiss of love campaign.



## OPEN ACCESS

**Citation:** Sankar CP, S. A, Kumar KS (2016) Learning from Bees: An Approach for Influence Maximization on Viral Campaigns. PLoS ONE 11 (12): e0168125. doi:10.1371/journal.pone.0168125

**Editor:** Peter Csermely, Semmelweis University, HUNGARY

**Received:** June 21, 2016

**Accepted:** November 9, 2016

**Published:** December 19, 2016

**Copyright:** © 2016 Sankar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data used in the work can be downloaded with suitable data mining tool from <http://www.twitter.com>.

**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Social networking platforms, such as Facebook and Twitter, have been extensively used for socio-political movements apart from viral marketing campaigns in the recent past as an increasing number of people spend more time online. In this scenario, it has become a challenge for the campaigners to diffuse the information quickly across the network. The influence maximisation problem originally introduced in the context of viral marketing is NP-hard for obtaining an optimal subset of users who can maximise the information diffusion [1–3]. It has resulted in an abundant development of approximate algorithms to identify prominent actors or the so-called *super spreaders* [4–6]. Study of such collective intelligence of individual users in socio-political campaigns will expand our capability to judge the human social behaviour and explain the influence diffusion process. Campaigns starting with people, who are the most influential individuals in society (also called initiators, influencers, key players, opinion leaders etc.) would lead to the maximum possible diffusion in minimum time. Hence, it is crucial for any campaign to select a small set of influencing seed users who have a strong influence on others. In this work, we propose a bio-inspired approach to select nodes in the initial set which can lead to a more rapid information propagation process and to simulate the diffusion by identifying retweet process of Twitter campaigns with the waggle dance of a bee colony.

Wu *et al.* have studied the spread of information in social groups and identified information relevant to one person is more likely to be of interest to individuals in the same social circle than those outside of it [7]. Information can be quickly spread between individuals within the micro-blogs in the form of word-of-mouth communication. Jansen *et al.* have examined Twitter as a medium for electronic word-of-mouth advertising and discuss the implications for corporates using micro-blogging as part of their overall marketing strategy [8]. Twitter hashtag adoption is a unique form of folksonomy since the initial adopters of the hashtag can be viewed as innovators and they attract or influence another group of users, namely, imitators [9]. Meeyoung and Haddadi compared three measures of influence: in-degree, retweets, and mentions and hypothesised that number of followers might not be a good measure of influence [10]. Besides, influence is not gained spontaneously or accidentally, but through concerted effort such as limiting tweets to a single topic. Weng and Jianshu proposed a topic-sensitive homophile based Page-rank measure for identifying influential users in Twitter and their *TwitterRank algorithm* measures the influence taking both the topical similarity between users and the link structure into account [11]. Gonzalez-Bailon *et al.* identified four types of users—namely, influential, hidden influential, broadcasters and common users—which can help the understanding of how users behave in diffusion processes [12]. Ramasuri and Narahari have applied cooperative game theory concept *shapely value* for solving the influence maximisation problem [13].

Nature was a source of inspiration for developing alternate algorithms for solving many real-world optimisation problems [14]. Bio-inspired algorithms, a major class of such algorithms, are popular mainly due to their efficiency. Swarm intelligence (SI), a branch of artificial intelligence (AI), is concerned with the design of intelligent multi-agent systems by taking inspiration from the collective behaviour of social insect colonies and other animal societies interacting locally with one another and with their environment by applying the concept of decentralised control and self-organization [15].

One of the major highlights in swarm intelligence domain is the development of ant algorithms for discrete optimisation in 1999 [16]. In 2003 modelling based on a natural behaviour of social insects was used for transportation problems [17]. Rivero *et al.* applied a biologically-inspired modification of ant colony optimisation algorithm for path search in social networks [18]. They extended the standard ACO algorithm by equipping the ants with a sense to smell similar to ant's natural pheromone-tracking capabilities. The extension adds further ability to follow the trail of food odour to the source. Bojic *et al.* have demonstrated recently how firefly synchronisation can be utilised for clustering of networks as well as data diffusion in machine networks [19]. Gao *et al.* proposed bio-inspired methodology, combining *physarum centrality* and *K-shell index*, to identify the influential node in a weighted network and compared its performance with other methods [20]. In the same year, an improved *Particle Swarm Optimisation (PSO)* algorithm was proposed by Zhang *et al.* to find the most influential users of Sina Weibo, the popular microblogging service in China [21]. Their algorithm utilises social interaction pattern to find an optimal solution. In Sina Weibo, users participate in network interaction by publishing tweets and retweets. The retweeting behaviour can be described as a variable of a user influence space which contains user experiences and surrounding network. Nikolaev *et al.* have recently introduced a metric, *viz., engagement capacity*, as a measure of the users to engage peers, to characterise online forum user behaviour and analyse the reach maximisation of online social media platforms [22]. Achananuparp *et al.* have taken the number of retweets as the engagement of Twitter user [23].

Among many swarm intelligence algorithms, Artificial Bee Colony (ABC) is the one which has been most widely studied and applied to solve the real world problems. Bees algorithm, introduced in 2005 by Dervis [24], is an optimisation algorithm inspired by the natural

foraging strategy of honey bees. Each candidate solution is considered as a food source (flower), and a population (colony) of agents (bees) is used to search the solution space. An artificial bee visits a flower (lands on a solution) and evaluates its profitability (fitness). The Artificial Bee Colony (ABC) algorithm introduced by Karaboga *et al.* in 2007 follows a new meta-heuristic approach inspired by the natural foraging behaviour of a honey bee to find the optimal food resources through a *waggle dance* [24–26]. A colony of honey bees constantly search the environment looking for new flower patches. Bees that found a highly profitable food source go to an area in the hive called the *dance floor*, and perform *waggle dance*. Through the waggle dance, a scout bee communicates the location of its discovery to other bees, which join in the exploration of the flower patch. Since the length of the dance is proportional to the scouts rating of the food source, more bees get recruited to harvest the best-rated flower patches. In 2009, Singh extended artificial bee colony algorithm for solving a constrained optimisation problem in minimum spanning tree [27]. In 2011, Pan *et al.* presented a discrete version of ABC for the lot-streaming flow shop scheduling problem [28]. Omkar *et al.* introduced a generic model based on the ABC for multi-objective design optimisation [29] in the same year. Akay and Karaboga introduced modified versions of ABC algorithm and applied them for efficiently solving real-parameter optimisation problems [30]. The modification is often related to the ratio of variance operator and the frequency of perturbation. Wu *et al.* presented an improved ABC algorithm to enhance the global search ability of basic ABC [31]. ABC algorithm was later upgraded for constrained optimisation problems [32]. Xu *et al.* described a chaotic ABC approach and applied to path planning of uninhabited combat air vehicle (UCAV) in various combat fields [33]. Zhang *et al.* modified ABC algorithm by changing the stages of *employed* and *onlooker* bees to promote the convergence rate [34]. Bio-inspired algorithms have many advantages such as the ability to provide multiple solutions, suitability of implementation in a parallel computing environment in addition to their capability of overcoming many shortcomings of traditional algorithms [35]. However, there are hardly few bio-inspired algorithms dealing with approximation problems in social network scenarios, the influence maximisation problem in particular.

In this paper, a new approach is proposed for simulation of influence propagation in viral campaigns through online micro-blogging platform Twitter for selecting a set of seed nodes which lead to maximum diffusion of influence in a particular context combining with a new node ranking procedure. Our approach is inspired by *waggle dance*, a communication process of honey bees, and utilises the global-local search capacity of the ABC algorithm to solve the influence maximisation problem. The effectiveness of the method is tested employing a retweet network of the hashtag #KissOfLove associated with one of the first socio-political movements in India which used Twitter very effectively.

## Proposed Method

It is extremely tough to identify the influential users during a collective action in micro-blogs without considering the features of tweets and user profiles. The proposed method focuses on approximating the optimal solution of influence maximisation problem using principles of swarm intelligence. The information available for each user is based on its activities and the knowledge of other individuals in the neighbourhood. Influence maximisation problem can be addressed by analysing the user, message, network and temporal features. This work combines all these features by analysing the tweet corpus and introduces a node ranking process by applying the concept of social interactions to find the optimal solution.

The ABC algorithm mimics the collective foraging behaviour of honey bees on searching food source. The central component of an ABC algorithm is combination global-local search

feature and waggle dance process. Through information exchanging and learning mechanism, the whole colony would always find relatively prominent food source. The ABC algorithm can be suitably adapted to gain useful knowledge from Twitter networks. Key personalities in social media who perform a *waggle dance*, by retweeting a particular tweet to attract their followers. We consider retweet of a particular tweet by a key user as a waggle dance. These key personalities may perform a *waggle dance* to attract their followers to a particular campaign. A user retweets a particular tweet to express his or her interests, beliefs, thoughts and concerns. So the social influence of Twitter profile and the favourability of tweet motivates one to retweet the tweets that help to represent their common interests. Such messages will be propagated within the social network through the connections between people, creating a good campaign that is highly personalised and engaging the interests of such users can almost be guaranteed to go viral.

The selection of the initial set of nodes from the network to start with the diffusion process play a crucial role in the influence maximisation problem. In general, the initial set is selected randomly. However, in this work we propose a bio-inspired approach for the selection of the initial set of  $k$  nodes which is significant in rapid convergence towards a sub-optimal solution in minimal runtime. Our approach, rather than selecting random  $k$ -seed nodes, introduces a node ranking process to rank the nodes by combining user profile features and message activities in the context. The proposed approach is divided into two parts: in the first part each node is ranked according to social reputation value and in second part top  $k$  nodes are selected as the initial  $k$ -node set. The diffusion process simulated with ABC algorithm detailed below.

## Node Ranking Process

The reputation of social profiles is the engine which powers any social media campaign. Profile features are used to identify the initial influence of each user and the influence value is updated based on tweet and retweet relationships. Finally, the reputation of each user is calculated based on total social activities in particular context. This social reputation rank is used to select the initial  $k$ -seed set. It is also used to select nodes for further exploration in decreasing order of their reputation rank in the network as detailed in Algorithm 1.

### Algorithm 1: Procedure for finding reputation rank of nodes

```

Data: Tweet corpus with a particular hashtag
Result: Reputation rank of each node
1 List all nodes in constructed digraph;
2: InitializeInfluenceValue(); ▷ for calculating profile influence value;
3: while Tweet corpus is non empty do
4:   if Message is a retweet then
5:     UpdateReTweet();
6:   else
7:     UpdateTweet();
8:   end
9:   UpdateInfluenceValue();
10: end
11: UpdateReputationRank();

```

The `InitializeInfluenceValue()` function is used to initialize the profile influence value of all nodes of the constructed digraph. Initially, each node has a slight influence value ( $Inf_0$ ) based on profile parameters such as the number of followers and number of users followed, shortly, *following*. Every user should have a sufficient number of followers to propagate his influence through tweets. However, more followers do not necessarily mean

more influence. The initial influence ( $Inf_0$ ), of each user, is calculated according to

$$Inf_0 = \frac{\text{Number of Followers}}{\text{Number of Following}} \tag{1}$$

and is assigned as the reputation of individual profiles. The celebrities and newsgroups profiles may have a huge number of followers, but they may follow very less number of people and hence  $Inf_0$  can take large values in the initial stage. Cha *et al.* have demonstrated that having a large number of followers does not contribute much to the influence of a user in the Twitter world whereas the number of users who actively retweet counts more [10]. Therefore, to reduce the significance of the number of followers in  $Inf_0$ , we introduce a normalisation process with parameters, namely, *Critical Value (CV)* and *Threshold Value (TV)*. The values of these parameters can be fixed based on the variation of  $Inf_0$ . The new normalised influence value is calculated as follows

$$Inf'_0 = \frac{Inf_0}{CV} \tag{2}$$

If  $Inf'_0$  value is greater than the *threshold value (TV)* then it is normalized again with *TV* using

$$Inf''_0 = TV + \frac{Inf'_0}{AveInf'_0} \tag{3}$$

where  $AveInf'_0$  is the average of  $Inf'_0$  over all nodes. After calculating the initial influence, each message in the dataset is explored to update the influence value by using `UpdateTweet ()` and `UpdateRetweet ()` functions which updates the tweet and retweet count respectively of the corresponding author node. The retweets received indicates the influence of a user. People retweet what they are interested. The more a tweet is retweeted, the more influence that tweet gets. Therefore, users who have more forwarded tweets can be considered influential users in normal situations.

If the user A retweets a tweet of user B, then the influence of user B is incremented:

$$Inf(B)_{t+1} = Inf(B)_t + Inf(A)_0. \tag{4}$$

This process is repeated until every message is processed. The final reputation rank of a user X in the social context is calculated iteratively:

$$\begin{aligned} Rep_t(X) = & Inf_t(X) - Inf_0(X) \\ & + \frac{\text{Number of Tweets}(X)}{\text{Total number of Tweets}} \\ & + \frac{\text{Number of re-tweets received}}{(\text{Number of re-tweets send} + 1)} \end{aligned} \tag{5}$$

### ABC Algorithm for influence Maximization

The starting point of any viral campaign on Twitter is the process of adopting a new hashtag. The small set of users recommend a particular hashtag in their tweets, their followers adopt it and retweet messages with the adopted hashtag, and it eventually becomes a widely recognised hashtag depending on the number of users adopting it.

**Algorithm 2:** Proposed algorithm for influence maximization in Twitter campaigns.

```

Data: Retweet network with reputation rank in each node
Result: K-Seed Nodes in hashtag context
1: Initialize  $E \subset V$  of employer bees with top  $k$  rank nodes;
2: Evaluate fitness value of each employer bee in  $E$ ;
3: Initialize scout bees in  $S$  with the nearest neighbour nodes of employer
   bees in  $E$ ;
4: Initialize onlooker bees in  $O$ ;
5: while termination condition is not met do
6:   for each scout bees  $i$  in  $S$  ▷ starting with highest Rank;
7:   do
8:     Calculate local fitness value of  $i$ ;
9:     if local fitness value of  $i >$  local fitness value of any  $j \in E$  then
10:      update the status of  $i$  as employer bee;
11:      update the status of  $j$  as onlooker bee;
12:      Update the onlooker bee status;
13:     end
14:   end
15:   Update scout bee;
16: end
17: Calculate global fitness value;
18: Return the set of  $k$  employer bees;

```

A retweet network  $G = (V, E)$ , shows tweet—retweet interactions associated with a particular hashtag within a group of individuals. It plays a fundamental role as a medium for the spread of information or influence among its members where  $V$  denotes the set of individuals (or nodes) and  $E$  denotes the set of edges. We consider a directed edge  $(u, v) \in E$  for any two nodes  $u, v \in V$  if  $v$  has retweeted any tweet of node  $u$ . In a typical influence maximisation scenario, one is interested to find  $k$  influencers to start with so that diffusion of influence is maximised. In the case of twitter campaigns the ultimate goal is to find the most engaging set of users from existing context who can be the initial set of adopters who lead to the most number of adoptions in similar socio-political campaigns in future. The analysis done by Cha *et al.* [10] have demonstrated that the most influential users can hold significant influence over a variety of themes. Here the problem is to choose a small set of influencing seed users who can get maximum influence on others. In many situations of collective action, only a few individuals among a group, for example, in the case of forage or travel, who have complete information, such as knowledge about the location of a food source, or of a migration route. Targeting these smaller proportion of informed individuals can achieve maximum diffusion of information within the entire network. The information diffusion happens through the social interactions that take place locally among the nodes that have direct connections.

We identify the analogy between individuals interacting on Twitter and bees perform waggle dance in a bee colony. The dynamic personalities in social media attract their followers to a particular campaign. We consider tweet—retweet activities of influencing user as a waggle dance which motivates us to adapt ABC algorithm to explore the retweet network to find the key influential nodes based on waggle dance. Each node in the given social context is considered as *flower patch*. The *employer bees*, who are used to locate influential opinion leaders in the network, are initially assigned to top  $k$  nodes from the node ranking process explained earlier. The *scout bees* are used to explore the nearest neighbour nodes of employer bees for better solutions. The *onlooker bees* indicate the followers (influenced) of influential opinion leaders. During the diffusion process, they are assigned with the status *influenced*. In each iteration the *local fitness value* is calculated by the maximum number of unique nodes that are influenced



by a single node. The *global fitness value* is maximum unique influenced nodes by a group of  $k$  nodes.

Input to Algorithm 2 is the retweet network  $G = (V, E)$  with node reputation rank value  $r_i$  of each node  $i$  and a positive integer  $k$  and the output is a subset  $A \subseteq V$  with  $|A| \leq k$  such that influence  $A$  is maximum. The proposed algorithm accept this input and initialize subset  $E \subset V$  of *employer bees* with nodes with top  $k$  ranks as the initial solution. Evaluate the local fitness value of each bee in  $E$  by counting the nodes that can be reached within  $k'$  steps and update the onlooker bee status of these nodes to be *influenced*. Store these influenced nodes in set  $O \subset V$  of *onlooker bees*. In this process, if a node is already influenced by another employer bee then it is avoided. The set  $S \subset V$  of *scout bees* is initialized with the nearest neighbour nodes of initial employer bees. Now we conduct the local search for optimal solutions to the problem by searching the neighbouring nodes in each iteration. An iterative process then starts, with selecting a scout bee with the highest reputation rank from  $S$ . If the local fitness value of a scout bee is greater than that of an employer bee then the employer bee is replaced with the respective scout bee. The replacement process updates the onlooker and scout bee status of the bees. In next step we insert the nodes with distance two from an employer bee is inserted into current scout bee list and repeat above steps until every node is influenced or the number of scout bees becomes null. The iterative process of the fitness evaluation ends when some termination condition is met, such as exceeding the execution time limit or a certain ratio of the nodes being influenced. The result, which is the best  $k$  node set for influence maximisation, is then returned. The final set of employer bees is identified as the set of  $k$ -influential nodes in the network which can maximise diffusion in the given context.

## Results and Discussion

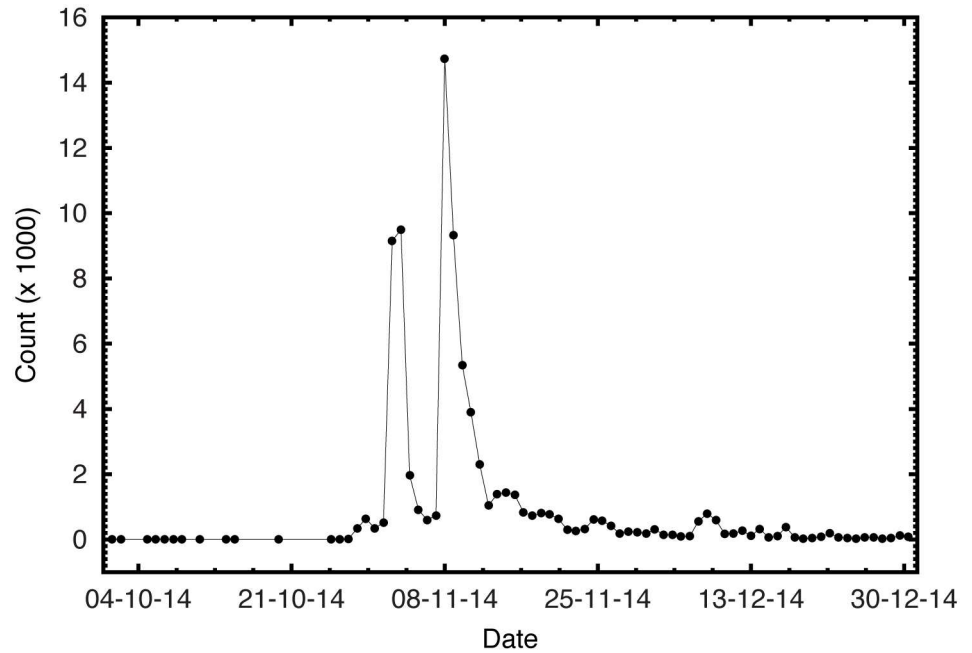
In this section, we report the results of the experiments with the proposed algorithm on twitter data associated with a hashtag #KissofLove. *Kiss of Love* protest is a non-violent protest against moral policing which started in Kerala and later spread to other parts of India. The movement began when a Facebook page named *Kiss of love* asked the youths across Kerala to assemble and kiss publicly as a protest against moral policing on November 2, 2014, at Kochi. The movement received widespread support in social media. After the initial protest in Kochi, similar protests were organised in the other main cities across the country. The protest was very popular on social networking sites and news media. Supporters of the campaign have been posting pictures of them kissing each other on social networking sites. Here we collected all tweets on online micro-blog Twitter under the hashtag #kissoflove and analysed to identify the most influential users.

We collected 78,735 tweets from 52,570 users from Twitter with #kissoflove using Socioviz tool dated between October 1, 2014 and December 31, 2014 (*c.f.* Table 1) complied with the terms of service. The plot of the number of tweets per day in Fig 1 shows two breakouts. The first breakout happened during the first protest in Marine Drive, Kochi on 2 November 2014 with 9194 tweets and the second breakout occurred during the protest in Delhi on November 8 with 14371 tweets. Fig 2 shows a typical example of the diffusion of a single tweet in short

**Table 1. The details of the retweet network of #KissofLove protest.**

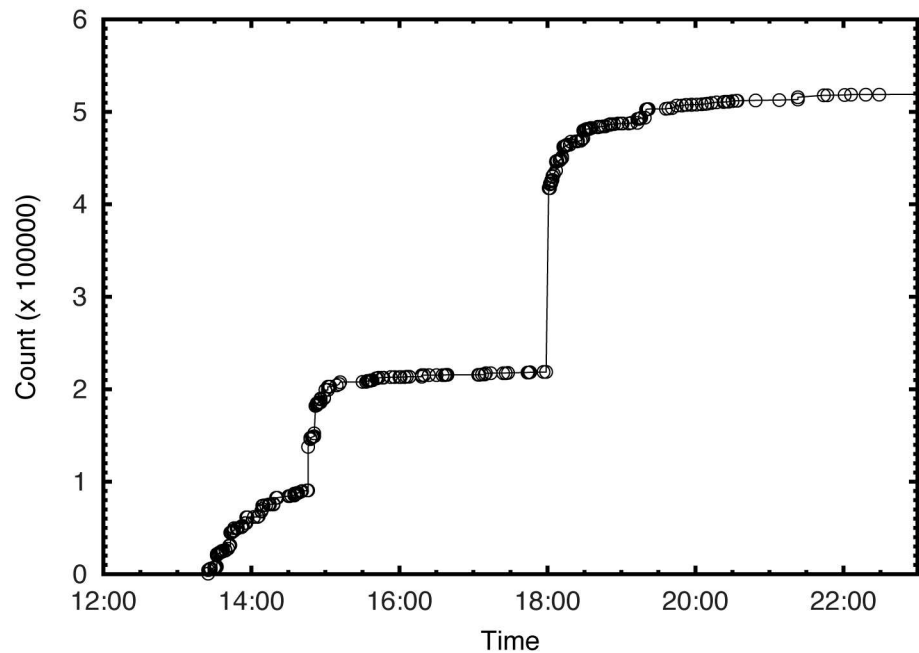
Network	Nodes	Edges
Complete network	22738	36082
Giant component	16855	35357

doi:10.1371/journal.pone.0168125.t001



**Fig 1. Growth of tweets in #KissofLove hashtag.**

doi:10.1371/journal.pone.0168125.g001



**Fig 2. The reach of a typical tweet on 2 November 2014.** The re-tweet at 18:01 hours by an influential user with 198871 followers causes a sudden jump in the reach.

doi:10.1371/journal.pone.0168125.g002



**Table 2. Structural properties of giant component.**

Structural Property	Value
Average Degree	2.01
Network Diameter	19
Avg Clustering Coefficient	0.029
Average Path Length	6.382

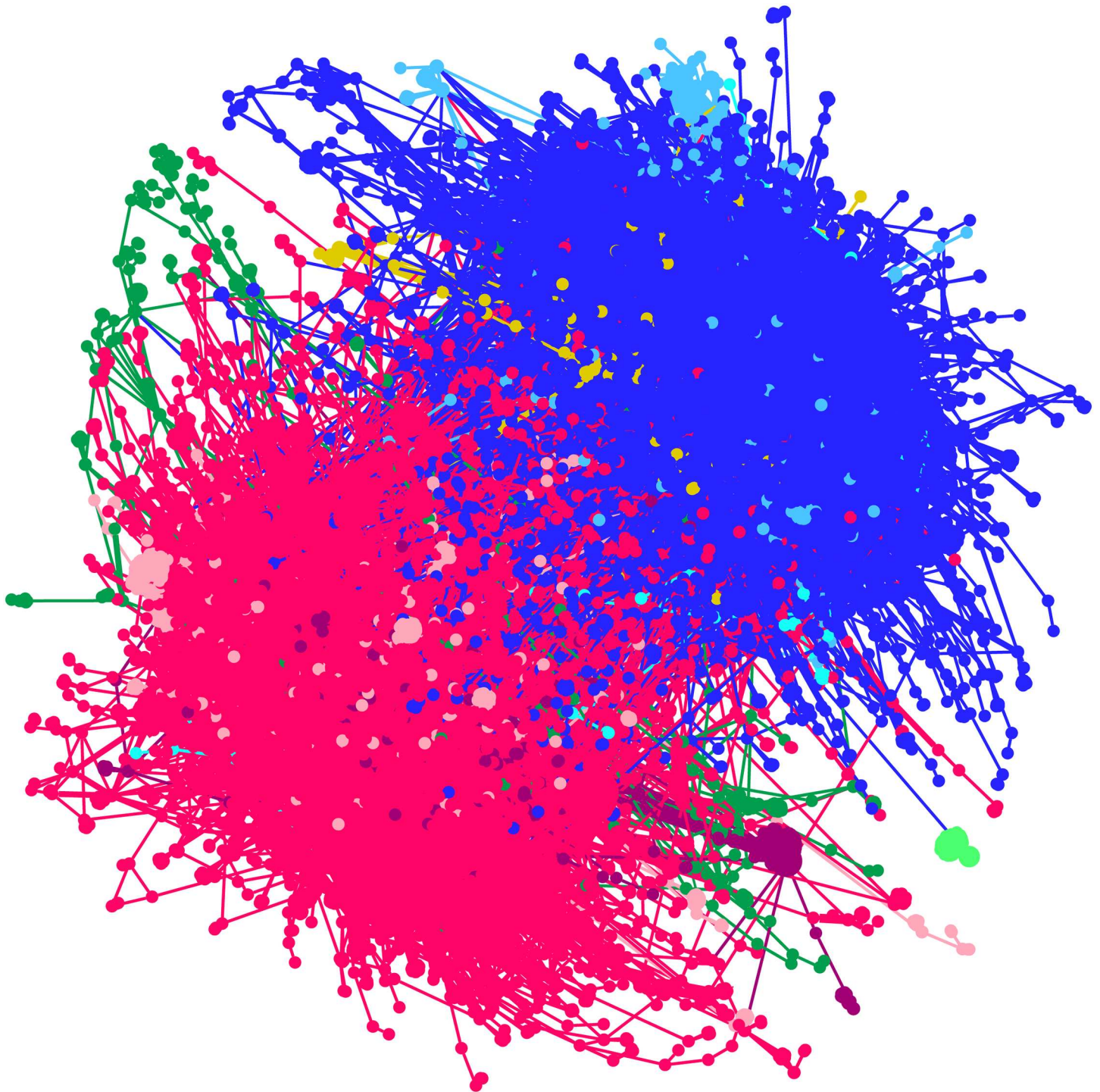
doi:10.1371/journal.pone.0168125.t002

span of time demonstrating the influencing role of re-tweets by the opinion leaders. The original tweet was retweeted by 350 times. However, a retweet by the user with 198871 followers at 18:01 hours causes a huge jump showing the impact of an influential user in the diffusion process.

A retweet network was constructed as a directed graph containing 22738 nodes representing all the active profiles and 36082 edges representing the re-tweeting relationships between these nodes. The network has a giant component with 16855 nodes and 35357 edges and 5168 islands. The largest among the islands has only 67 nodes. We considered the giant component for further analysis. As a first step, we computed the structural properties of giant component and is given in [Table 2](#)

The giant component was further subjected to cluster analysis [36, 37] with resolution parameter 2.0. A cluster in a network is a group of similar nodes. Here the objective is to understand how opinion on this movement is divided. The main clusters are shown in [Fig 3](#). Content analysis of the tweets was carried out to understand the general characteristics of each cluster. The cluster of nodes in blue colour represents supporters (43.02%) and the cluster in red colour are opponents (55.99%), and this composition is interesting in a sociological perspective. The remaining clusters are not significant.

As the first step in the implementation of proposed algorithm, a set of top 10 nodes (identified by ID) of reputation value ( $Rep_t$ ), without normalisation procedure given in Eqs (2) and (3), is selected and is given in [Table 3](#). It is evident from the table that media houses and celebrities have got high influence because of their large number of followers whereas they are following less number of people. As noted earlier the number of follower does not contribute much to influence diffusion in terms of retweets. As these users seldom retweet they were avoided from initial set by the normalisation procedure discussed previously and set both the *critical value (CV)* and the *threshold value (TV)* to be 10 for the implementation of the algorithm. The normalisation process is introduced to provide priority to individual users other than media houses. Top 10 users according to final reputation rank are shown in [Table 4](#). It may be noted that some media houses still retains its position regarding reputation value showing their potential to contribute to the diffusion process. The reputation value indicates the worthiness of a specific user within a specific context or campaign. With the selected seed set of 10 nodes, the influence propagation is simulated according to Algorithm 2. Local search or exploration process analyses the neighbour nodes of a seed node and if they have higher fitness value than that of the seed nodes then update the existing list of seed nodes with new nodes. This procedure simplifies the computation by selecting initial nodes by the node ranking process to obtain a feasible estimate. The proposed procedure combines the fitness value obtained from local and global exploration to find the optimal solutions. This procedure is repeated until every node is explored. Here the algorithm returns final ranks of the users base on  $Rep_t$ . The list of nodes with top ten ranks as given [Table 5](#).



**Fig 3. Clusters of re-tweet network.** The two major clusters in blue and red colours represent supporters and opposers respectively.

doi:10.1371/journal.pone.0168125.g003

**Table 3. Top ten profiles with reputation value ( $Rep_t$ ) without normalisation.**

Rank	ID	Followers	Following	Tweets	Retweets	Retweets Received	$Rep_t$	Profile Type
1	18716	144481	1	6	0	265	246.8541	Cricket Celebrity
2	3380	8395159	61	2	0	69	235.6157	News
3	13235	108723	0	1	0	11	188.2344	News
4	9325	227290	4	1	0	21	103.1516	Fashion Celebrity
5	1100	429436	12	2	0	14	68.6661	News
6	20506	1750558	52	2	0	24	65.1878	News
7	6118	526340	28	0	1	5	40.8162	News
8	13226	11691	0	2	0	4	29.1656	News
9	21054	10665	0	16	0	51	27.4836	Social blogger
10	4015	10252	0	1	0	8	26.8066	No details

doi:10.1371/journal.pone.0168125.t003

**Table 4. Initial top ten users identified by node ranking process based on  $Rep_t$  as per Algorithm 1 after normalisation.**

Rank	ID	Initial Reputation ( $Inf_0$ )	Tweets	Retweets	Re-tweets received	$Rep_t$	Profile Type
1	12595	10.07094	40	80	840	3690.502	Social Blogger
2	22245	5.1356	21	110	1188	2712.612	Social Blogger
3	11626	10.70384	12	0	792	1448.124	News
4	10790	10.18415	65	0	233	1361.254	Social Blogger
5	19973	5.8342	18	0	496	1302.437	Journalist
6	6654	10.03122	15	66	332	1210.094	No details
7	16324	5.833	19	2	825	1128.391	Social Blogger
8	3379	10.12924	8	2	367	1013.984	News
9	2907	10.01723	16	15	255	850.2413	Social Blogger
10	11322	1.3535	32	37	342	817.9727	Social Blogger

doi:10.1371/journal.pone.0168125.t004

**Table 5. Final top ten users identified by Algorithm 2 with maximum influence in the network.**

Rank	ID	Initial Reputation ( $Inf_0$ )	Tweets	Retweets	Re-tweets received	$Rep_t$	Profile Type
1	12595	10.07094	40	80	840	3690.502	Social Blogger
2	22245	5.1356	21	110	1188	2712.612	Social Blogger
3	15428	13.483	18	2	345	264.12	Social Blogger
4	21054	27.48361	38	10	322	495.482	RTI Activist
5	19973	5.8342	18	0	496	1302.437	Journalist
6	2907	10.01723	16	15	255	850.2413	Social Blogger
7	2	15.44	12	8	225	594.391	Social Blogger
8	9325	103.12	48	4	123	413.984	Social Blogger
9	11322	1.3535	32	37	342	817.9727	Social Blogger
10	6654	10.03122	15	66	332	1210.094	No details

doi:10.1371/journal.pone.0168125.t005

**Table 6. Comparison of the rate of diffusion starting with seed nodes selected by different ranking methods.** The numbers under each iteration show the number of nodes reached with respective ranking measure used for the selection of initial set.

SI No	Ranking measure for seed node selection of initial set	Iteration number								
		1	2	3	4	5	6	7	8	9
1	Rep <sub>t</sub> (as per Table 5)	3683	6758	8537	9429	9929	9982	10017	10018	10018
2	Rep <sub>t</sub> (as per Table 4)	3503	6581	8623	9645	9898	9936	9970	9971	9971
3	Out-degree centrality	3246	6823	8729	9613	9815	9855	9877	9890	9890
4	Degree centrality	3497	6360	8438	9315	9784	9844	9865	9866	9866
5	Inf <sub>0</sub> '	596	1561	3703	6195	7572	8301	8742	8778	8793
6	No of tweets	998	1340	4471	6749	7709	8153	8361	8417	8425
7	Betweenness centrality	1617	4140	6537	7738	8275	8349	8381	8389	8403
8	No of retweets	1220	3637	6294	7630	8231	8332	8380	8389	8403
9	In-degree centrality	786	2735	5701	7433	8182	8323	8354	8382	8389
10	Closness centrality	25	40	45	50	54	73	95	396	1157

doi:10.1371/journal.pone.0168125.t006

To check the effectiveness of the proposed node ranking procedure, we have compared the selection of initial seed nodes based on our approach with standard ranking approaches for selecting the k-seed nodes based on different measures. We constructed the initial sets based on ten ranking measures, viz., centralities such as degree, betweenness, closeness, activities such as tweeting and retweeting, and reputations such as Rep<sub>t</sub>. Starting with each initial seed node set, information diffusion was simulated using the giant component and computed the number of nodes influenced up to several iterations. The reach of influence with various initial sets of seed nodes is given in Table 6 in which the numbers shown in the cells are the number of nodes reached (informed) up to the iteration level starting with the set of initial nodes selected by corresponding ranking measure. The advantage of seed node selection based on Rep<sub>t</sub>, over all other measures, is clearly evident from the first iteration which is expected to improve the diffusion process through this initial set of adopters. After the first iteration, the diffusion could reach (the number of nodes informed) to 3683 nodes by initial seed node set selected according to Rep<sub>t</sub> which is higher than that of the initial set selected by all other methods. It may be noted that values in the second row, obtained by initial set selected from taking the output of Algorithm 1, is closer to the values in the first row obtained by the initial set selected from the output of Algorithm 2. This shows that the seed nodes selected by the node ranking procedure introduced in this paper, considering social interaction in terms of tweets and retweets, is much closer to the optimum solution. The degree centrality does not reflect the global characteristics of a network whereas betweenness and closeness centralities do not consider the location of nodes [20]. Results of our simulations given in Table 6 show that out-degree and degree centralities perform comparatively better than betweenness and closeness centralities though they are inferior to Rep<sub>t</sub>. The results also show that the number of tweets or retweets alone does not add value to a user profile for the diffusion process compared to other measures. The proposed procedure based on Rep<sub>t</sub> has got the highest influence function value in the 7th iteration indicating a faster and wider diffusion of influence and hence more nodes may be influenced in subsequent iterations. It is evident from the table that the proposed reputation-based ranking method is superior for the rate of diffusion compared to other centrality measures based ranking. The influential nodes identified according to the bio-inspired approach introduced in this paper can be useful in future socio-political Twitter campaigns as Cha *et al.* [10] have observed that most influential users can hold significant influence over a variety of topics.

## Conclusion

The quick diffusion of socio-political campaigns through online micro-blogging platforms such as Twitter is extremely dependent on identification seed nodes. In this paper, a new node ranking method is proposed and compared, with indices popularly used for the same, indicating its advantage. The complexity of the exact influence maximisation problem has attracted many approximation algorithms in the literature of which bio-inspired algorithms are hardly few. We propose a bio-inspired algorithm identifying the retweet process with waggles of the bee colony. The performance is assessed with the retweet network formed by the #KissofLove campaign on Twitter. Results of the experiment show that the algorithm combined with the proposed node ranking method can effectively identify the opinion leaders online Twitter campaign. Besides, we also report results of the exploratory analysis of the retweet network of the #KissofLove protest.

## Acknowledgments

The authors would like to thank Mr Alexandro Zonin for providing the dataset of the #KissofLove campaign and his SocioViz web tool. The authors also wish to thank campus computing facility of the University of Kerala set up under DST-PURSE programme for providing computational facilities.

## Author Contributions

**Conceptualization:** CPS AS KSK.

**Data curation:** CPS.

**Formal analysis:** CPS AS KSK.

**Investigation:** CPS AS KSK.

**Methodology:** CPS.

**Project administration:** AS KSK.

**Resources:** AS KSK.

**Software:** CPS.

**Supervision:** AS KSK.

**Validation:** CPS AS KSK.

**Visualization:** CPS AS KSK.

**Writing – original draft:** CPS AS KSK.

**Writing – review & editing:** CPS AS KSK.

## References

1. Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2002. p. 61–70.
2. Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2003. p. 137–146.
3. Kempe D, Kleinberg J, Tardos É. Influential nodes in a diffusion model for social networks. In: Automata, languages and programming. Springer; 2005. p. 1127–1138.



4. Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. *Theory of Computing*. 2015; 11(4):105–147. doi: [10.4086/toc.2015.v011a004](https://doi.org/10.4086/toc.2015.v011a004)
5. Pei S, Muchnik L, Andrade JS Jr, Zheng Z, Makse HA. Searching for superspreaders of information in real-world social media. *Scientific reports*. 2014; 4. doi: [10.1038/srep05547](https://doi.org/10.1038/srep05547)
6. Zhu Y, Wu W, Bi Y, Wu L, Jiang Y, Xu W. Better approximation algorithms for influence maximization in online social networks. *Journal of Combinatorial Optimization*. 2015; 30(1):97–108. doi: [10.1007/s10878-013-9635-7](https://doi.org/10.1007/s10878-013-9635-7)
7. Wu F, Huberman BA, Adamic LA, Tyler JR. Information flow in social groups. *Physica A: Statistical Mechanics and its Applications*. 2004; 337(1):327–335. doi: [10.1016/j.physa.2004.01.030](https://doi.org/10.1016/j.physa.2004.01.030)
8. Jansen BJ, Zhang M, Sobel K, Chowdury A. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*. 2009; 60(11):2169–2188. doi: [10.1002/asi.21149](https://doi.org/10.1002/asi.21149)
9. Rogers EM. A prospective and retrospective look at the diffusion model. *Journal of Health Communication*. 2004; 9(S1):13–19. doi: [10.1080/10810730490271449](https://doi.org/10.1080/10810730490271449) PMID: [14960401](https://pubmed.ncbi.nlm.nih.gov/14960401/)
10. Cha M, Haddadi H, Benevenuto F, Gummadi PK. Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM*. 2010; 10:10–17.
11. Weng J, Lim EP, Jiang J, He Q. Twitterrank: finding topic-sensitive influential twitterers. In: *Proceedings of the third ACM international conference on Web search and data mining*. ACM; 2010. p. 261–270.
12. González-Bailón S, Borge-Holthoefer J, Moreno Y. Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist*. 2013;p. 0002764213479371.
13. Narayanam R, Narahari Y. A shapley value-based approach to discover influential nodes in social networks. *IEEE Transactions on Automation Science and Engineering*. 2011; 8(1):130–147. doi: [10.1109/TASE.2010.2052042](https://doi.org/10.1109/TASE.2010.2052042)
14. Fister Jr I, Yang XS, Fister I, Brest J, Fister D. A brief review of nature-inspired algorithms for optimization. *arXiv preprint arXiv:13074186*. 2013;.
15. Kennedy J, Kennedy JF, Eberhart RC, Shi Y. *Swarm intelligence*. Morgan Kaufmann; 2001.
16. Dorigo M, Caro G, Gambardella L. Ant algorithms for discrete optimization. *Artificial life*. 1999; 5(2):137–172. doi: [10.1162/106454699568728](https://doi.org/10.1162/106454699568728) PMID: [10633574](https://pubmed.ncbi.nlm.nih.gov/10633574/)
17. Teodorovic D. Transport modeling by multi-agent systems: a swarm intelligence approach. *Transportation Planning and Technology*. 2003; 26(4):289–312. doi: [10.1080/0308106032000154593](https://doi.org/10.1080/0308106032000154593)
18. Rivero J, Cuadra D, Calle J, Isasi P. Using the ACO algorithm for path searches in social networks. *Applied Intelligence*. 2012; 36(4):899–917. doi: [10.1007/s10489-011-0304-1](https://doi.org/10.1007/s10489-011-0304-1)
19. Bojic I, Lipic T, Podobnik V. Bio-inspired clustering and data diffusion in machine social networks. In: *Computational Social Networks*. Springer; 2012. p. 51–79. doi: [10.1007/978-1-4471-4054-2\\_3](https://doi.org/10.1007/978-1-4471-4054-2_3)
20. Gao C, Lan X, Zhang X, Deng Y. A bio-inspired methodology of identifying influential nodes in complex networks. *PloS one*. 2013; 8(6):e66732. doi: [10.1371/journal.pone.0066732](https://doi.org/10.1371/journal.pone.0066732) PMID: [23799129](https://pubmed.ncbi.nlm.nih.gov/23799129/)
21. Zhang B, Zhong S, Wen K, Li R, Gu X. Finding high-influence microblog users with an improved PSO algorithm. *International Journal of Modelling, Identification and Control*. 2013; 18(4):349–356. doi: [10.1504/IJMIC.2013.053540](https://doi.org/10.1504/IJMIC.2013.053540)
22. Nikolaev A, Gore S, Govindaraju V. Engagement Capacity and Engaging Team Formation for Reach Maximization of Online Social Media Platforms. In: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD'16. New York, NY, USA: ACM; 2016. p. 225–234. Available from: <http://doi.acm.org/10.1145/2939672.2939681>
23. Achananuparp P, Lim EP, Jiang J, Hoang TA. Who is retweeting the tweeters? modeling, originating, and promoting behaviors in the twitter network. *ACM Transactions on Management Information Systems (TMIS)*. 2012; 3(3):13.
24. Karaboga D. An idea based on honey bee swarm for numerical optimization. Technical report-tr06, Erciyes university, engineering faculty, computer engineering department; 2005.
25. Karaboga D, Basturk B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*. 2007; 39(3):459–471. doi: [10.1007/s10898-007-9149-x](https://doi.org/10.1007/s10898-007-9149-x)
26. Karaboga D, Akay B. A comparative study of artificial bee colony algorithm. *Applied Mathematics and Computation*. 2009; 214(1):108–132. doi: [10.1016/j.amc.2009.03.090](https://doi.org/10.1016/j.amc.2009.03.090)
27. Singh A. An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem. *Applied Soft Computing*. 2009; 9(2):625–631. doi: [10.1016/j.asoc.2008.09.001](https://doi.org/10.1016/j.asoc.2008.09.001)
28. Pan QK, Tasgetiren MF, Suganthan PN, Chua TJ. A discrete artificial bee colony algorithm for the lot-streaming flow shop scheduling problem. *Information sciences*. 2011; 181(12):2455–2468. doi: [10.1016/j.ins.2009.12.025](https://doi.org/10.1016/j.ins.2009.12.025)



29. Omkar S, Senthilnath J, Khandelwal R, Naik GN, Gopalakrishnan S. Artificial Bee Colony (ABC) for multi-objective design optimization of composite structures. *Applied Soft Computing*. 2011; 11(1):489–499. doi: [10.1016/j.asoc.2009.12.008](https://doi.org/10.1016/j.asoc.2009.12.008)
30. Akay B, Karaboga D. A modified artificial bee colony algorithm for real-parameter optimization. *Information Sciences*. 2012; 192:120–142. doi: [10.1016/j.ins.2010.07.015](https://doi.org/10.1016/j.ins.2010.07.015)
31. Wu XJ, Hao D, Xu C. An improved method of artificial bee colony algorithm. In: *Applied Mechanics and Materials*. vol. 101. Trans Tech Publ; 2012. p. 315–319.
32. Brajevic I, Tuba M. An upgraded artificial bee colony (ABC) algorithm for constrained optimization problems. *Journal of Intelligent Manufacturing*. 2013; 24(4):729–740. doi: [10.1007/s10845-011-0621-6](https://doi.org/10.1007/s10845-011-0621-6)
33. Xu C, Duan H, Liu F. Chaotic artificial bee colony approach to Uninhabited Combat Air Vehicle (UCAV) path planning. *Aerospace Science and Technology*. 2010; 14(8):535–541. doi: [10.1016/j.ast.2010.04.008](https://doi.org/10.1016/j.ast.2010.04.008)
34. Zhang X, Zhang X, Ho S, Fu W. A modification of artificial bee colony algorithm applied to loudspeaker design problem. *Magnetics, IEEE Transactions on*. 2014; 50(2):737–740. doi: [10.1109/TMAG.2013.2281818](https://doi.org/10.1109/TMAG.2013.2281818)
35. Halim AH, Ismail I. Bio-Inspired optimization method: A review. *NGGT Journal: International Journal of Information Systems*. 2014; 1:1–6.
36. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*. 2008; 2008(10):P10008. doi: [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008)
37. Lambiotte R, Delvenne JC, Barahona M. Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:08121770*. 2008;.