Artificial Bee Colony-Based Influence Maximization Approach for Social Coding Portal GitHub

Anuja Arora, Jaypee Institute of Information Technology, Noida, India
Riyu Bana, Jaypee Institute of Information Technology, Meerut, India
Habib Shah, King Khalid University, Abha, Saudi Arabia
Divakar Yadav, Madan Mohan Malaviya University of Technology, Gorakhpur, India

ABSTRACT

Influence maximization is the main source of virality of any social media post/marketing activity. In recent trends, influence maximization has moved towards analytic approach instead of just being a suggestive metaphor for various social media paradigm. In this article, ego-centric approach and a bio-inspired algorithm is applied on social coding community, Github, for influence maximization. First, developers’ and projects’ egocentric network-based studies are conducted to find out influential developer and project based on varying social media measures. Second, artificial bee colony (ABC) bio-inspired algorithm is used to select social bees (i.e., developers set and projects set in rapid convergence towards an optimal solution to achieve influence maximization). Algorithm result ensures the best solution in terms of social community connectivity path optimization in less run time while finding most influential social bees.

KEYWORDS
Artificial Bee Colony, Bio-Inspired Approach, Github, Influence Maximisation, Network-Based Analysis, Social Coding Community, Social Network Analysis, Virality

1. INTRODUCTION

Software development is exploiting social media and transforming software development with an astonishing speed in recent years. Various social coding platforms have unfolded sharing of code as social coding pattern such as Github, Bitbucket and Stack Overflow etc. In this research work, Github has been chosen as platform for code influence maximization problem and to identify developers and project those maximally influence to the coding crowd. Github is viewed as one of an unprecedented social coding network which has pulled in countless developers in a very short timeframe. Developers from all round the world utilize it to save their public and private projects furthermore to team up online with different developers to contribute in new projects. The world of software engineering is being tested with the rapid increase in the number of developers and repositories on Github.
GitHub is considered as collection of codes and contributors for varying programming languages and technologies. A comprehensive study in the area of Github network analysis highlights many research gaps, some of them are as follows:

- In the previous studies, a few properties of the system structure of social coding in Github were estimated. For the future work, a recommendation system could be created to pick appropriate developers to cooperate for specific tasks in GitHub. Information of most influential projects and developers could also be of concern for recruiters. So, plausibility of prescribing suitable candidates to different recruiters could be explored;
- Social behaviour standard about fork network, pull-request network and watch-network of GitHub could be considered in future. In light of these social standards of conduct, a recommendation system could be outlined which can push the most important projects to clients.

Therefore, this social coding network is used to handle the code and developer influence maximization problem. Influence maximization is primarily introduced in the context of viral marketing. Influence maximization is used to find out subset of users those can maximize the information influence. Similarly, to obtain an optimal subset of developers and repositories those maximizes the information influence. To identify most influential set of developers and repositories for influence maximisation Ant colony optimization (ABC) and bio-inspired algorithm has been used. Bio-Inspired algorithms form a compendium of biology inspired computing and these algorithms have foundation based on living things’ behaviour. Few works have been done in order to apply bio-inspired algorithm for influence maximization in social networks. Various bee swarm intelligence (SI) algorithms exist under bio-inspired algorithm. Some are Yang proposed virtual bee algorithm (2005), Wedde, Farooq & Zhang, (2004) presented presented Bee–Hive algorithm, Bee Colony Optimization is discussed by (Teodorovic and Dellorco, 2005), etc. Out of all existing Bee SI algorithms, Ant Bee colony (ABC) optimization has been most widely used algorithm, which was introduced by Karaboga et. al. in 2007. ABC is inspired by natural forging behaviour of honey bee and engages in a Meta heuristic activity to discover the optimal food resources via a waggle dance.

In this work, ABC has been applied to find out optimal solution to achieve influence maximization of developer and repository towards an optimal solution. Basically, In this study, work is performed in two directions:

- **Ego-centric network based interpretation:** Various networks have been created using networkx python library and visualisation is done using Gephi software. Project-project and developer-developer network is used to identify influential set of developers and projects for influence maximization;
- **Identification of influential set of developers and repositories:** Using ant colony optimization for influence maximization.

Paper structure is as follows- related work is discussed in section II, experiments which are performed to find out influencer index are detailed in section III and IV. Section III discusses network creation and visualisation approaches and Section IV discusses about Bio-inspired approach. Corresponding results are enlisted in section V and finally paper is concluded in section VI.

### 2. RELATED WORK

Social coding platform have not just pulled in the developers from all round the world but has likewise impacted the consideration of researchers towards it. Many researchers have studied the Github network from varied perspectives to analyze different properties.
Towards a Trust Management Enabled Identity Metasystem
Weiliang Zhao and Jian Yang (2012). Intelligent and Knowledge-Based Computing for Business and Organizational Advancements (pp. 118-134).
www.igi-global.com/chapter/towards-trust-management-enabled-identity/65790?camid=4v1a

Microwave Circuit Design
Gabriel Cormier and Tyler Ross (2013). Swarm Intelligence for Electric and Electronic Engineering (pp. 18-39).
www.igi-global.com/chapter/microwave-circuit-design/72821?camid=4v1a