

Realistic influence maximization based on followers score and engagement grade on instagram

Kristo Radion Purba¹, Yulia²

¹School of Computer Science and Engineering, Taylor's University, Malaysia

²Informatics Department, Petra Christian University, Indonesia

Article Info

Article history:

Received May 24, 2020

Revised Nov 13, 2020

Accepted Feb 17, 2021

Keywords:

Diffusion model

Influence maximization

Information diffusion

Instagram

Social network

ABSTRACT

In recent years, the emergence of social media influencers attracts the study of a realistic influence maximization (IM) technique. The theoretical performance of IM has become matured. However, it is not enough since IM has to be implemented in a social media environment. Realistic IM algorithms and diffusion models have been proposed, such as the addition of user factors or a learning agent. However, most studies still relied on the influence spread benchmark, which makes the usefulness questionable. This research is among the first IM study using Instagram data. In this study, two diffusion models are proposed, which are based on the original IC and LT models, with the addition of the engagement grade (EG) factor. An algorithm called IMFS (IM with followers score) is proposed to accommodate the new models as well as IC and LT. In addition, realistic benchmark methods are proposed, namely the average engagement of the activated users, and the overlapping between post likers and activated users. The result shows that the proposed models are 2-3x more realistic if compared to IC and LT.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Kristo Radion Purba

School of Computing and IT, Taylor's University

1, Jalan Taylors, 47500 Subang Jaya, Selangor, Malaysia

Email: kristoradionpurba@sd.taylors.edu.my

1. INTRODUCTION

Social networks are growing at an unprecedented rate, and it has opened various business opportunities, especially brand marketing. Compared to other social networks, Instagram is the best platform to target millennial audience [1], and the platform with the highest engagement [2]. Choosing influencers can be a difficult task, where common beliefs such as picking influencers with the highest number of followers and likes don't always produce the best results [3].

Choosing influencers that can spread influence to a maximum number of audience with a minimum budget is widely studied in a field called influence maximization (IM) [4]. An IM algorithm is used to generate seeds set (influencers) that produces the best possible influence spread (the number of activated users) under specific diffusion models [5]. However, the commonly used diffusion models, i.e., linear threshold (LT) and independent cascade (IC), assume that each user has a similar level of influence degree and susceptibility. This makes LT and IC models less useful in the real world, even though some state-of-the-art IM algorithms such as IMM [6] and SSA [7] can produce a very high influence spread.

The number of influence spread itself has been matured, where recent theoretical improvements are only in terms of runtime [8]. Recent studies are more focused on making IM more realistic. Incorporation of various factors have been studied, such as influence susceptibility [9], sentiment [10, 11], freeloaders [12], targeted ads [13], and engagement [14, 15]. There are also bandit-based IM algorithms [16, 17], which

typically uses feedbacks from actual data. However, the benchmark methods were still based on influence spread, which makes the usefulness in real-world questionable.

This study aims to develop diffusion models and an IM algorithm that activate more engaging users. Two new diffusion models based on LT and IC by incorporating engagement value are proposed, namely IC-eg and LT-eg (EG=engagement grade). An IM algorithm called IMFS (influence maximization with followers score) is proposed to provide the best solution for the proposed diffusion models. In addition, realistic and practical benchmark methods are proposed, based on the average engagement rate and engagement grade of the activated users, and the overlap between the activated users and actual post likers. To best of our knowledge, this is the first study of IM using Instagram data.

The following questions are studied in this research, i.e. (1) does incorporation of engagement grade produce a more realistic influence maximization? (2) how realistic is the proposed diffusion models if compared to the classic IC and LT models? This study is a step towards a practical IM, that can be used by business users to choose brand marketers more realistically. The rest of this paper is organized in the following sections, i.e., related studies, methodology, experimental results, conclusion.

2. RELATED STUDIES

There were recent studies on improving the theoretical and real-world performance of IM. Influence spread and runtime are commonly used as the theoretical benchmarks. The first notable state-of-the-art IM algorithm was TIM and TIM+ algorithms [18], with remarkably high influence spread and low runtime. The influence spread was further improved by IMM [6] by adding martingale. Further improvement in runtime was made by SSA [7], which was up to 1,200 faster than IMM. More recently, machine learning-based IM algorithms emerged, such as DISCO [8], which improved the runtime of SSA. However, DISCO required a training phase that took up to three days of execution. Thus far, IMM and SSA are the best performers in terms of influence spread, while DISCO is considered to have similar performance.

Real-world improvement on IM can be made through bandit agents or incorporation of factors. In a bandit-based IM, IM is executed multiple times, and the algorithm tunes the cumulative regret parameter based on the outcome of the diffusion process [19]. Unlike usual IM, the target of bandit-based IM is to minimize cumulative regret, i.e., the loss of influenced nodes cumulated from every iteration. However, the vital part of bandit-based IM, i.e. the outcome of diffusion, was mostly synthesized using techniques such as graph sampling [16, 20] and diffusion random vector [19].

The influence factor has been studied [9]; however, the study used a prediction technique called correlated label propagation (CLP) to generate influence degree and susceptibility values. These values were predicted based on a synthetic and real graph instead of actual values. Engagement is one of the most prevalent factors in IM, with engagement forms such as conversation content and reply [14], assortativity, influence on second neighborhoods [15], network topology [21], silent users [22]. However, these studies were either relied on assumptions [15] or only worked in a limited environment [14]. Furthermore, the influence spread remains to be the most popular benchmark method, which remains theoretical.

3. METHODOLOGY

This section discusses the data preparation, engagement rate (ER) and engagement grade (EG) metrics, the proposed IM diffusion models and algorithm. The proposed IM algorithm was tuned to work with the existing IC and LT models, as well as the proposed models.

3.1. Data preparation

The dataset used in this study was collected from Instagram on April to May 2020 from the followers of 24 private universities in Malaysia. This localization was intended to create many connections among users. From the users, the related data was collected, i.e., posts, hashtags, post liker, and followers. This was done using Instagram API and various third-party Instagram websites. The users were cleaned using the fake user's classification model from an earlier study [23]. The raw data consists of 70,409 nodes/users, 1,007,107 edges/connections, 1,031,348 posts, and 47,689,496 likers entry. The simplified and anonymized user data and the network are available at <https://www.kaggle.com/krpurba/im-instagram-70k-eg>. There are two network data, i.e., a network for IC and LT models, and a network for IC-eg and LT-eg.

3.2. Engagement rate and engagement grade metrics

Engagement rate (ER) is among the most popular metrics for social networks, which is defined by the number of (likes+comments) divided by followers divided by the number of posts. Comparing users with ER, however, is not fair across users with different size of followers, where a high number of followers leads

to lower ER [24]. Based on the average ER across a different number of followers [24], we established engagement grade (EG), which ranged from 0.0 to 1.0. The average ER and followers are shown in Table 1.

Table 1. Average ER across size of followers

Tier	1	2	3	4	5	6	7	8	9	10	11
Followers min	0	2k	5k	10k	25k	50k	75k	100k	150k	250k	500k
Followers max	2k	5k	10k	25k	50k	75k	100k	150k	250k	500k	1m
Average ER %	10.7	6.0	4.9	3.6	3.1	2.6	2.5	2.5	2.4	2.9	2.3

Engagement grade (EG) is formulated (1).

$$EG(user) = \min \left(\frac{ER(user)}{3 \times ER \text{ baseline (followers)}}, 1.0 \right) \quad (1)$$

where: ER baseline=The average ER for the user's number of followers

For any user, EG value between 0.0 to 0.33 is below average ER, between 0.33 to 0.67 is above average ER, and between 0.67 to 1.0 is far above average. EG value gives a fair reward for users across the different size of followers. For example, the same EG=2.0 will be assigned to users with (ER=4.6 and followers=800 k) and (ER=21.4 and followers=1 k). EG value is capped at 1.0.

3.3. IM diffusion models

This study proposes two diffusion models, namely IC-eg (independent cascade-EG) and LT-eg (linear threshold-EG). Compared to the respective original models, these models only modify the edge weight by adding a multiplication of EG, as formulated below:

$$P_u(user, flr) = 2 \times EG(user) \times \frac{1}{followers(flr)} \quad (2)$$

where: P_u is edge probability for IC-eg model (or edge weight for LT-eg model) from a user to a follower

The constant of 2 can be set to an arbitrary number to keep a reasonable influence spread value. In our dataset, the average EG of all users is 0.395. Removing this constant causes a very low influence spread. The addition of engagement in the edge weight was due to its usefulness in indicating several things, i.e. (1) Low engagement means a possibly high number of fake followers [25]. On FakeCheck.co (<https://www.fakecheck.co>), a website for fake followers analysis, the ER of a user is compared to "industry standard", which most likely means it uses an EG-like metric instead of ER, (2) Using influencers with high engagement rate leads to more effective marketing [26], (3) High engagement means high activity (less passive users) [27], (4) Engagement tells naturally means how much a user is liked in the society.

3.4. Influence maximization with followers score algorithm

The proposed IMFS algorithm mainly focused on the IC-eg and LT-eg models. However, in the experiments, this algorithm also worked well with IC and LT models. The main basic idea of IMFG comes from "calculating the followers in multiple depths using a sampled graph." A sampled graph was commonly used in IM algorithms with RR-Set (reverse-reachable set) [6, 18], which is created by removing each edge with a probability of (1- edge weight).

The RR-Set collects influential users by generating the sampled graph several times and keeping the users who frequently appear in the graph [7]. In contrast, this research generates sampled graph only one time. The graph is used to calculate *followers score*, which is the aggregation of the number of followers in multiple depths (up to 10), which is formulated (3).

$$flrs(user) = \sum_{depth=0}^9 2^{-depth} \times flr(depth, user) \quad (3)$$

where: $flr(depth, user)$ =The number of followers at *depth*, where *depth*=0 means direct followers

IMFS algorithm uses a sampled graph to minimize runtime, which means only users in the sampled graph are calculated. The whole process of IMFS is shown in Figure 1. IMFS starts with *followers score* calculation (*estimation phase*). During *snum* (number of seeds)=1, the *flrs* of all users are calculated. In *snum*>1, however, *flrs* calculation is stopped when there is no improvement in the last *noimpr_flrs* loops.

Before designing the next phases of the algorithm, we further examined the usefulness of *flrs* in terms of directly predicting *influence spread*. By simulating all users individually (*snum*=1), it was found that

flrs has a correlation (Pearson's) of 0.86 to influence spread in the IC model, and 0.84 in IC-eg model. Since these numbers are not extremely high, some inaccuracies are expected to happen during the "conversion" of *flrs* to *influence spread*. Thus, IMFS requires *simulation phase* that simulates several candidates.

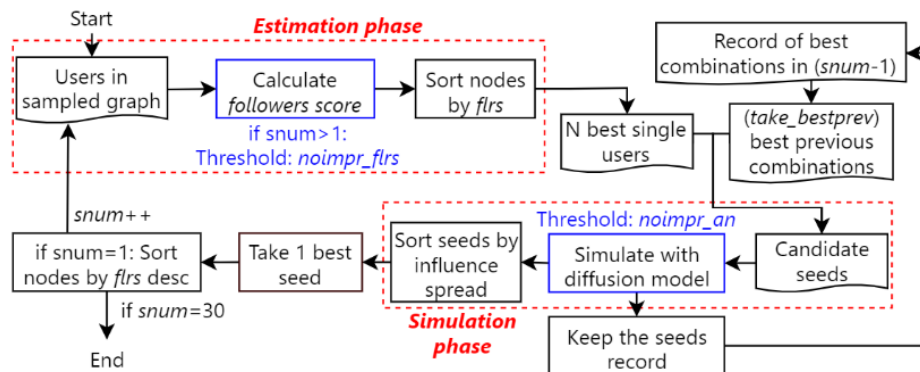


Figure 1. IMFS algorithm

To mitigate the potential inaccuracies during the "conversion", we added three parameters, i.e. *noimpr_flrs*, *noimpr_an*, *take_bestprev*. Removing *noimpr_flrs* and *noimpr_an* simply means executing a greedy algorithm, which sacrifices runtime. The *take_bestprev*, on the other hand, aims to extend the classic greedy, which only uses one best previous combination. The final parameter values that were used in this research are Figure 2 *noimpr_flrs*=50, *noimpr_an*=50, *take_bestprev*=5. The parameters have effects on the influence spread result. The values were acquired by increasing the values gradually until breaking points of influence spread were reached, as seen in Figure 2. Note that these experiments were done by adjusting a single parameter at a time while keeping the others at the default values.

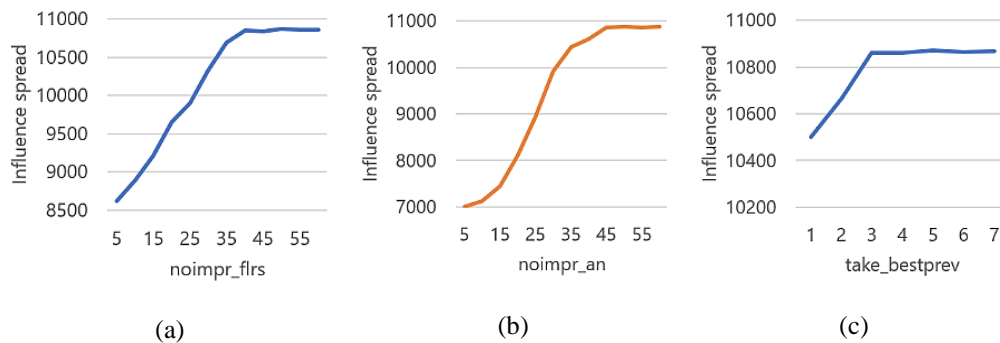


Figure 2. Effect of increasing parameters value on influence spread in *snum*=30, under IC-eg model, (a) *noimpr_flrs*, (b) *noimpr_an*, (c) *take_bestprev*

In the simulation phase, each new candidate is combined with *take_bestprev* best previous combinations and simulated with $\epsilon=0.1$. This ϵ value was used by many studies [6, 28] to get an accurate enough influence spread. Since during the estimation phase, users are already sorted by *flrs* in descending order, users with higher *flrs* are expected to have higher influence spread (with some inaccuracies). If there is no improvement in the last *noimpr_an* candidates, the simulation phase is stopped, and the best candidate will be taken as the current seed. The pseudocode of IMFS is Algorithm 1.

Algorithm 1. IMFS algorithm

```

//----Estimation Phase----
allcandidates = []
foreach bestPrevSeeds as bpseed, limit take_bestprev{
  candidates = []
  foreach allnodes as node{
    cand = merge(bpseed,node)
    cand.flrs = calculate_flrs(diffusionmodel,cand)
    candidates.push(cand)
    if no_improvement(candidates[flrs],noimpr_flrs) and snum>1: break
  }
  allcandidates = merge(allcandidates,cand)
}
sort allcandidates by flrs descending
//----Simulation Phase----
candidatesFinal = []
foreach allcandidates as cand{
  cand.inf_spread = simulate(cand,diffusionmodel)
  candidatesFinal.push(cand)
  if no_improvement(candidatesFinal[inf_spread],noimpr_an): break
}
sort candidatesFinal by inf_spread descending
bestPrevSeeds = candidatesFinal
finalseed = candidatesFinal[0]

```

4. EXPERIMENTAL RESULTS

As the baseline algorithms, IMM [6] and SSA [7] were chosen since they are the best performers in terms of influence spread and runtime (SSA). There were four diffusion models to be tested, i.e., the classic IC and LT models and the IC-eg and LT-eg models. There were four benchmark methods, i.e., influence spread and runtime, with the addition of user metric benchmarks, i.e.:

a. Average engagement (EG and ER)

When influencers are simulated according to a diffusion model, a number of users are activated. The simulations were executed 1,000 times, and the average EG and ER of the activated users were calculated. Activating less engaging users means activating passive users, which is not realistic.

b. Likers overlap (LO)

This LO value assumes that a user is influenced if he/she liked an influencer's post, regardless of being a follower or outsider. An additional form of influence, such as product buyers, are much harder to get. The LO value is the proportion of activated users who are likers, which is formulated (4).

$$LO(\text{influencer}) = 100 \times \frac{\text{Likers} \cap \text{activated users}(\text{influencer})}{\text{Activated users}(\text{influencer})} \quad (4)$$

4.1. Synthetic benchmarks

The performance under IC and LT models are shown in Figure 3. The proposed IMFS algorithm performed similarly in terms of influence spread compared to other algorithms. The runtime of SSA is still much superior, which consistently runs under 1s, while IMFS is around 2-3x faster than IMM.

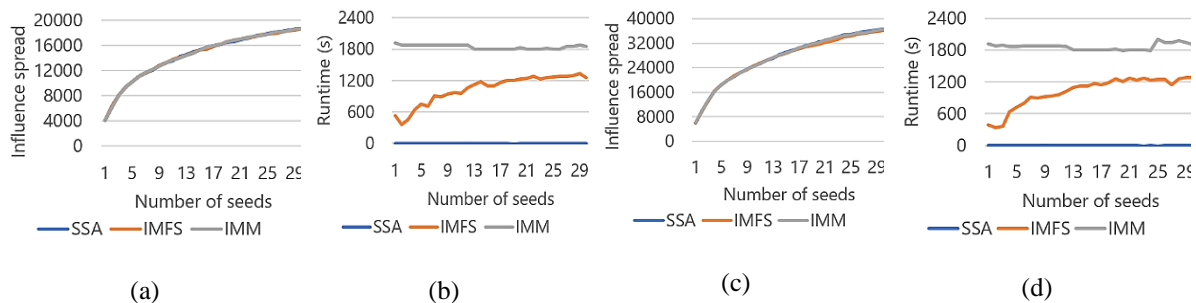


Figure 1. Performance under IC and LT models, (a) Inf. spread (IC), (b) Runtime (IC), (c) Inf. spread (LT), (d) Runtime (LT)

The existing IMM and SSA algorithms should work well for IC-eg and LT-eg models since only edge weights adjustment were made in the network. All algorithms performed similarly under IC-eg model, as shown in Figure 4. However, IMFS outperforms the influence spread of other algorithms under LT-eg model. This means that the use of followers score is more suitable for LT-eg, if compared to RR-set.

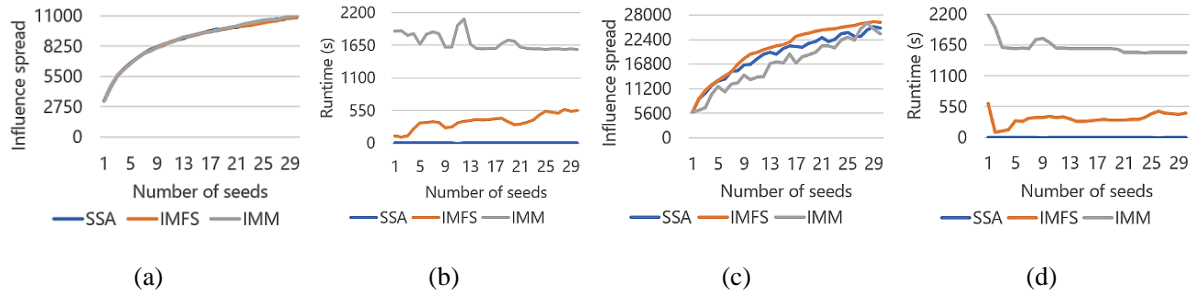


Figure 4. Performance under IC-eg and LT-eg models, (a) Inf. spread (IC-eg), (b) Runtime (IC-eg), (c) Inf. spread (LT-eg), (d) Runtime (LT-eg)

4.2. User metric benchmarks

Most algorithms performed almost similar in terms of the average EG and ER. However, the difference is between diffusion models. As can be seen in Figures 5, 6, and Table 2, LT-eg outperforms other models, with IC-eg as the second-best, while IC and LT models have much lower EG and ER. This shows that both LT-eg and IC-eg are more realistic. IMFS algorithm performed slightly better in EG and ER under IC-eg and LT-eg models if compared to other algorithms.

Table 2. Average EG and ER of activated users (snum=1 to 30)

Algorithm	Engagement Grade (EG)				Engagement Rate (ER)			
	IC	LT	IC-eg	LT-eg	IC	LT	IC-eg	LT-eg
IMFS	0.388	0.379	0.523	0.540	12.941	12.561	17.869	18.424
SSA	0.388	0.378	0.520	0.540	12.936	12.502	17.759	18.404
IMM	0.388	0.378	0.520	0.534	12.915	12.514	17.751	18.258

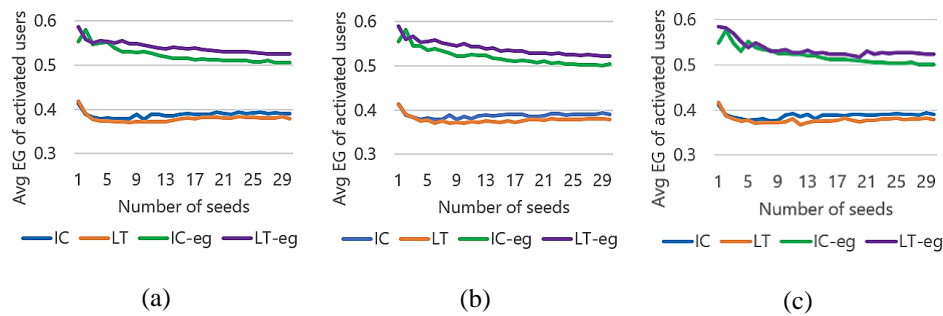


Figure 5. Average engagement grade (EG) of activated users, (a) IMFS, (b) SSA, (c) IMM

In terms of likers overlap (LO), IC-eg and LT-eg models performed 2-3x better than IC and LT models, as seen in Figure 7 and Table 3. IMM performed not as good as other algorithms in LT-eg, similar to the influence spread result. This shows that IMM is not suitable for LT-eg, while SSA can keep up. Based on the average LO provided in Table 3, IMFS performed better than other algorithms.

4.3. Seeds similarity

The outcome of an IM algorithm is the chosen seeds set (influencers). Practically, for example, a business user queries the algorithm to produce ten influencers for marketing purposes. Allocating budget for the influencers is a difficult task, where accurate identification of influencers become crucial. The following Figure 8 shows the seeds set similarity for each number of seeds, compared to IC-eg, as the most realistic

model (based on LO). The LT-eg results of IMFS and SSA are the closest resemblance to IC-eg. IMM algorithm continues to not perform well under LT-eg model. Using the IMFS algorithm as a representative, the overall average of seeds similarity are Figure 8 IC=51.88%, LT=50.77%, LT-eg=73.69%. It can be concluded that the final decision (the chosen influencers) of IC and LT models can be around 50% different with IC-eg (the most realistic model). This significant difference can have a big impact on business users.

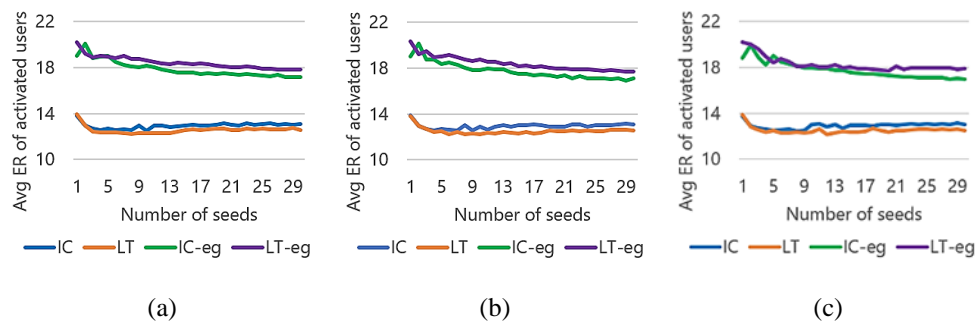


Figure 6. Average engagement rate (ER) of activated users, (a) IMFS, (b) SSA, (c) IMM

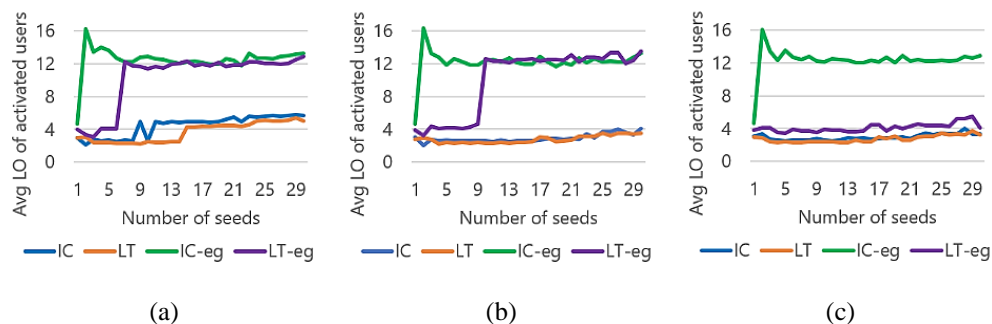


Figure 7. Average likers overlap (LO) of activated users, (a) IMFS, (b) SSA, (c) IMM

Table 3. Average LO of activated users (snum=1 to 30)

Algorithm	IC	LT	IC-eg	LT-eg
IMFS	4.437	3.666	12.489	10.326
SSA	2.918	2.774	12.196	10.051
IMM	2.961	2.748	12.324	4.110

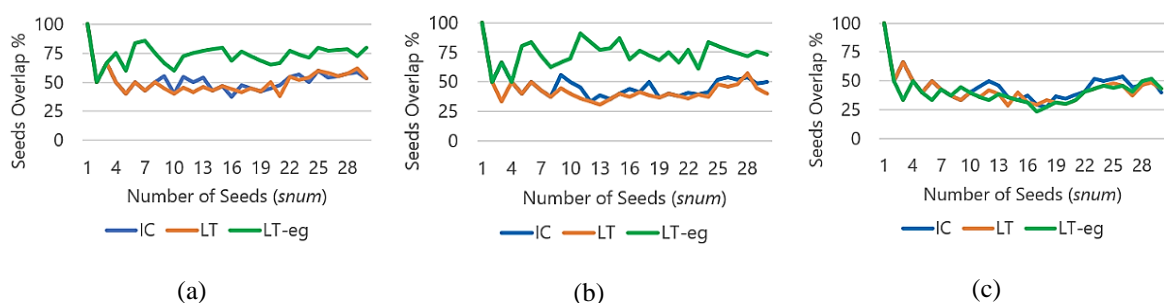


Figure 8. Seeds set similarity (overlap) compared to IC-eg model, (a) IMFS, (b) SSA, (c) IMM

5. CONCLUSION

Identifying influencers have become an important task in recent years. This problem is studied through influence maximization (IM) research. IM studies have become mature theoretically. Recent studies

were more focused on making IM more realistic by adding user factors, such as engagement, sentiment, multiple network analysis, or learning during the diffusion process.

This research added users' engagement value as the measure of activeness in the network, as well as proposed user-based benchmark methods. Experimental results showed that the proposed IC-eg and LT-eg diffusion models were superior in terms of the average EG, ER, and likers overlap (LO) of the activated users if compared to IC and LT. The high values of these user metrics have proven that the proposed models are more realistic, and can produce more engaging users, than IC and LT. The produced LO value by IC-eg and LT-eg is 2-3x better than IC and LT, indicates that the proposed models are closer to reality. The less realistic models (IC and LT) have a 50% difference in terms of the chosen influencers if compared to IC-eg.

The proposed IMFS algorithm, which was explicitly tuned for IC-eg and LT-eg, produced slightly better results in terms of the user metrics if compared to SSA and IMM. Furthermore, IMFS achieved better influence spread under LT-eg model if compared to other algorithms, while performing similarly under other models. This has proven that the followers score, which is the backbone of IMFS algorithm, is suitable for all diffusion models. In future work, additional user metrics can be added, such as followers growth. The edge weights of the diffusion models can also be further tuned to achieve higher user metrics. To enhance the practical usage, topic consideration has to be added to suit marketing on specific brand category.

REFERENCES

- [1] M. Iqbal, "Instagram Revenue and Usage Statistics (2019)," 6 November 2019. [Online]. Available: <https://www.businessofapps.com/data/instagram-statistics/>.
- [2] B. Feehan, "2019 Social Media Industry Benchmark Report," 15 February 2019. [Online]. Available: <https://www.rivaliq.com/blog/2019-social-media-benchmark-report/>.
- [3] M. D. Veirman, V. Cauberghe e L. Hudders, "Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude," *International Journal of Advertising*, vol. 36, no. 5, pp. 798-828, 2017.
- [4] D. Kempe, J. Kleinberg e É. Tardos, "Maximizing the spread of influence through a social network," *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137-146.
- [5] D. Dupuis, C. D. Mouza, N. Travers e G. Chareyron, "Real-Time Influence Maximization in a RTB Setting," *Data Science and Engineering*, vol. 5, no. 3, pp. 224-239, 2020.
- [6] Y. Tang, Y. Shi e X. Xiao, "Influence Maximization in Near-Linear Time: A Martingale Approach," *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 2015, pp. 1539-1554.
- [7] H. T. Nguyen, M. T. Thai e T. N. Dinh, "Stop-and-Stare: Optimal Sampling Algorithms for Viral Marketing in Billion-scale Networks," *Proceedings of the 2016 International Conference on Management of Data*, 2016, pp. 695-710.
- [8] H. Li, M. Xu, S. S. Bhowmick, C. Sun, Z. Jiang e J. Cui, "DISCO: Influence Maximization Meets Network Embedding and Deep Learning," *arXiv:1906.07378v1*, 2019.
- [9] S. Aral e P. S. Dhillon, "Social influence maximization under empirical influence models," *Nature Human Behaviour*, vol. 2, pp. 375-382, 2018.
- [10] Y. Li, W. Chen, Y. Wang e Z.-L. Zhang, "Influence Diffusion Dynamics and Influence Maximization in Social Networks with Friend and Foe Relationship," *Proceedings of the sixth ACM international conference on Web search and data mining*, 2013, pp. 657-666.
- [11] S. Jendoubi e A. Martin, "Evidential positive opinion influence measures for viral marketing," *Knowledge and Information Systems (KAIS)*, vol. 62, pp. 1037-1062, 2019.
- [12] Y. Chen, H. Li e Q. Qu, "Negative-Aware Influence Maximization on Social Networks," *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pp. 8063-8064, 2018.
- [13] J. Guo, P. Zhang, C. Zhou, Y. Cao e L. Guo, "Personalized influence maximization on social networks," *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 199-208.
- [14] X. Tang e C. C. Yang, "Ranking User Influence in Healthcare Social Media," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 3, no. 4, pp. 73-96, 2012.
- [15] P. D. Karampourniotis, B. K. Szymanski e G. Korniss, "Influence Maximization for Fixed Heterogeneous Thresholds," *Scientific Reports*, vol. 9, no. 5573, 2019.
- [16] S. Vaswani, L. V. Lakshmanan e M. Schmidt, "Influence Maximization with Bandits," *arXiv:1503.00024*, 2015.
- [17] Q. Wu, Z. Li, H. Wang, W. Chen e H. Wang, "Factorization Bandits for Online Influence Maximization," *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19)*, Anchorage, AK, USA, 2019.
- [18] Y. Tang, X. Xiao e Y. Shi, "Influence Maximization: Near-Optimal Time Complexity Meets Practical Efficiency," *Proceedings 2014 ACM SIGMOD International Conference on Management of Data*, 2014, pp. 75-86.
- [19] S. Vaswani, B. Kveton, Z. Wen, M. Ghavamzadeh, L. V. Lakshmanan e M. Schmidt, "Model-Independent Online Learning for Influence Maximization," *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, pp. 1-10, 2017.
- [20] S. Lei, S. Maniu, L. Mo, R. Cheng e P. Senellart, "Online Influence Maximization," *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, 2015, pp. 645-654.

- [21] A. Calio, R. Interdonato, C. Pulice e A. Tagarelli, "Topology-Driven Diversity for Targeted Influence Maximization with Application to User Engagement in Social Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 12, pp. 2421-2434, 2018.
- [22] R. Interdonato, C. Pulice e A. Tagarelli, "An Influence Maximization based approach to the Engagement of Silent Users in Online Social Networks," *Symposium on Advanced Database Systems (SEBD)*, 2017.
- [23] K. R. Purba, D. Asirvatham e R. K. Murugesan, "Classification of instagram fake users using supervised machine learning algorithms," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, pp. 2763-2772, 2020.
- [24] K. R. Purba, D. Asirvatham e R. K. Murugesan, "Instagram Post Popularity Trend Analysis and Prediction using Hashtag, Image Assessment, and User History Features," *International Arab Journal of Information Technology (IAJIT)*, vol. 18, no. 1, pp. 85-94, 2021.
- [25] Captiv8, "#NoFilter: How to Uncover the Fakes, 2018 Fraud Influencer Marketing Benchmark Report," *Captiv8*, 2018.
- [26] S. Yang, S. Lina, J. R. Carlson e W. T. Ross Jr., "Brand engagement on social media: will firms' social media efforts influence search engine advertising effectiveness?," *Journal of Marketing Management*, vol. 32, no. 5-6, pp. 526-557, 2016.
- [27] P. Bellavista, L. Foschini e N. Ghiselli, "Analysis of Growth Strategies in Social Media: the Instagram Use Case," *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Limassol, Cyprus, 2019, pp. 1-7.
- [28] M. Rossi, "Graph Mining for Influence Maximization in Social Networks," *Université Paris-Saclay*, 2017.

BIOGRAPHIES OF AUTHORS



Kristo Radion Purba is currently a PhD candidate in Computer Science at Taylor's University Malaysia. His main research interests are data analysis, social media, and machine learning. He is also an independent apps developer, data scientist, and lecturer. His programming expertise is web, mobile apps, and game development. He worked on several PC and mobile game titles from 2009 to 2013 as an independent developer. From 2013 to 2018, he worked as a contracted programmer at PT. HM. Sampoerna, Tbk (an affiliate of Philip Morris International). From 2014 to 2018, he worked as a lecturer of informatics at Petra Christian University. Email: kristoradion@live.com/kristoradionpurba@sd.taylors.edu.my



Yulia is an associate professor in Informatics Department, Petra Christian University, Surabaya, Indonesia. She received her bachelor degree in Informatics Department from Surabaya University, Indonesia, and master degree in Information Technology from University of Indonesia. Her research interest is enterprise resource planning. Email: yulia@petra.ac.id