

REVIEW

Open Access



Influence maximization on temporal networks: a review

Eric Yanchenko^{1*}, Tsuyoshi Murata² and Petter Holme^{3,4}

*Correspondence:
eyanchenko@aui.ac.jp

¹ Global Connectivity Program,
Akita International University,
Akita, Japan

² Department of Computer
Science, Tokyo Institute
of Technology, Tokyo, Japan

³ Department of Computer
Science, Aalto University, Espoo,
Finland

⁴ Center for Computational Social
Science, Kobe University, Kobe,
Japan

Abstract

Influence maximization (IM) is an important topic in network science where a small seed set is chosen to maximize the spread of influence on a network. Recently, this problem has attracted attention on temporal networks where the network structure changes with time. IM on such dynamically varying networks is the topic of this review. We first categorize methods into two main paradigms: single and multiple seeding. In single seeding, nodes activate at the beginning of the diffusion process, and most methods either efficiently estimate the influence spread and select nodes with a greedy algorithm, or use a node-ranking heuristic. Nodes activate at different time points in the multiple seeding problem, via either sequential seeding, maintenance seeding or node probing paradigms. Throughout this review, we give special attention to deploying these algorithms in practice while also discussing existing solutions for real-world applications. We conclude by sharing important future research directions and challenges.

Keywords: Graphs, Diffusion, Dynamic networks

Introduction

Networks, or graphs, are a simple tool to abstractly represent a system involving interacting entities, where the objects are modeled as nodes and their relationship as edges (Strogatz 2001; Newman 2003, 2018). Because of their generality and flexibility, many real-world settings have leveraged networks over the past few decades including: online social networks (Garton et al. 1997; Mislove et al. 2007; Phuvipadawat and Murata 2010), infrastructure networks (Latora and Marchiori 2005; Liu and Song 2020; Guimera and Amaral 2004) and biological process networks (Girvan and Newman 2002; Pavlopoulos et al. 2011). Recently, there has been great interest in not only understanding the topological structure of networks, but also how information diffuses on them (López-Pintado 2008; Rodriguez et al. 2011; Xu and Liu 2010; Harush and Barzel 2017). For example, in social networks, we may be interested in understanding viral outbreaks in a population, or breaking news spreads in an online setting.

The most fundamental assumption of network science and machine learning applied to graphs is that the network structure begets the function of the networked system. First discussed in abstract terms by Georg Simmel in the 1890s (Simmel 1955) and in

the language of graph theory by Jacob Moreno and Helen Jennings in the 1930s (Moreno and Jennings 1938), this assumption is close to the core structuralism—a pillar of 20th-century (primarily social) science. It suggests that we can infer the function of nodes from their position in the network.

The foundational functional concept is a node's importance. But to operationalize a concept like importance, we must consider many specifics about the system. Some questions may include: What is the objective of the system? What dynamics operate on it? Or what are the possible interventions? *Influence maximization (IM)*,¹ assumes a scenario where some *diffusion* (in the mathematical and physical literature, also known as *spreading*) process can happen on the network and we want this process to reach as many nodes as possible. This diffusion process is triggered by some seed nodes and the IM problem is to identify the seed nodes that maximize the number of nodes affected by the diffusion.

Viral (or word-of-mouth) marketing is an obvious potential application area that fits the assumptions above (Domingos and Richardson 2001; Leskovec et al. 2007; Hinz et al. 2011; Lü et al. 2016; Bhattacharya et al. 2019). Recommendation systems are another relevant area of application (Herlocker et al. 2004; Bobadilla et al. 2013; Aggarwal 2016; Zhang et al. 2021; Huang et al. 2022), as is seeding public health campaigns (Yadav et al. 2016, 2018; Wilder et al. 2017, 2018). However, by analogy to network centrality—another family of conceptualizations under the umbrella term of importance—*influence maximization* is interesting for a wider area of problems. Protecting critical infrastructure (Liu and Song 2020) or safeguarding against bioterrorism (Waniek et al. 2022) could also benefit from *influence maximization* studies. It is also a possible approximation (Holme 2017) for distinct but related scenarios like the vaccination problem (Holme 2004; Lee et al. 2012) (finding nodes whose removal would hinder a diffusion event as effectively as possible), and sentinel surveillance (Christakis and Fowler 2010; Bai et al. 2017) (identifying nodes that would be suitable probes for early and reliable detection of diffusion events).

Kempe et al. (2003) first formulated the IM problem in their seminal work and ever since, the problem has been explored extensively in the computer science, statistical physics and information science literature. The majority of work considers static networks where the nodes and edges are fixed (Kempe et al. 2003; Bharathi et al. 2007; Chen et al. 2009; Goyal et al. 2011). Most methods either efficiently emulate the influence spread function, or use some heuristic to rank nodes by importance. We eschew extended discussion of the static IM problem and instead refer interested readers to reviews in Li et al. (2018), Azaouzi et al. (2021). In many real-world situations, the static assumption is violated as networks have temporal variation with links forming and disappearing (Holme and Saramäki 2012; Holme 2015; Li et al. 2017). For example, in contact networks, a person's interactions vary greatly throughout a single day or week. Being at home, work, or a mall may lead to sizeable differences in both the frequency and set of contacts. Accounting for this dynamism is essential for effective IM seeding because, e.g., the temporal variation in the network structure greatly

¹ Throughout this work, we use the terms *influenced activated* and *infected* interchangeably when referring to a node's state.

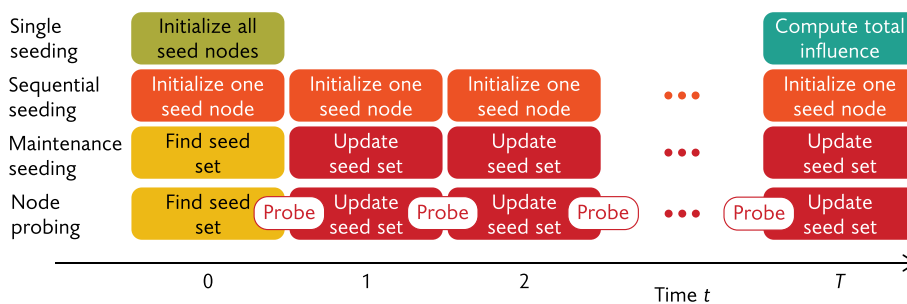


Fig. 1 Different temporal influence maximization paradigms compared in this work

impacts the rate and extent of the diffusion process (Prakash et al. 2010; Karsai et al. 2011). The topic of this review is IM techniques and analyses for such dynamically varying networks.

There are several key challenges associated with the IM problem. First, simply calculating the expected influence spread is #P-hard for many models (Goyal et al. 2011). One common approach to circumvent this issue is Monte Carlo (MC) simulations, where the diffusion process is simulated a large number of times and the average number of influence nodes is used to estimate the influence spread (Kempe et al. 2003; Ohsaka et al. 2014). Still, selecting the optimal seed nodes is NP-hard (Kempe et al. 2003). Therefore, many researchers employ heuristics so finding the globally optimal solution is rarely guaranteed. In temporal networks, there is an additional challenge stemming from the interplay of dynamic variation in edge sets and diffusion processes.

In this work, we provide a comprehensive review of the existing literature on IM on temporal networks and elucidate important future research areas. One main contribution of this review is our keen eye towards the challenges associated with deploying these methods in practice. While there has been significant research on the static and temporal IM problem, we found that there is minimal research on using these methods “in the field.” Thus, we highlight the utility of each method for practitioners. This differentiates the present work from reviews in Yang and Pei (2019) and Hafiene et al. (2020). Additionally, Yang and Pei (2019) focuses on influence analysis rather than strictly influence maximization. We also discuss several tasks not mentioned by the authors, e.g., sequential seeding where nodes are initialized throughout the process, and the *ex ante* setting where the future evolution of the network is unknown.

There are four main sections of this review. For the remainder of this section, we introduce the necessary prerequisites for studying the IM problem. In section “Single seeding”, we discuss “single seeding” methods which select a single seed set at the beginning of the diffusion process. This problem is the natural extension of the static IM problem. We classify the existing methods into three categories while also discussing methods which analyze the diffusion process. The topic of section “Multiple seedings” is methods which repeatedly choose seed nodes as the network evolves. Within this category, some methods activate nodes at different times throughout the diffusion process while others “maintain” an influential seed set. Additionally, we consider the node probing problem where the future evolution of

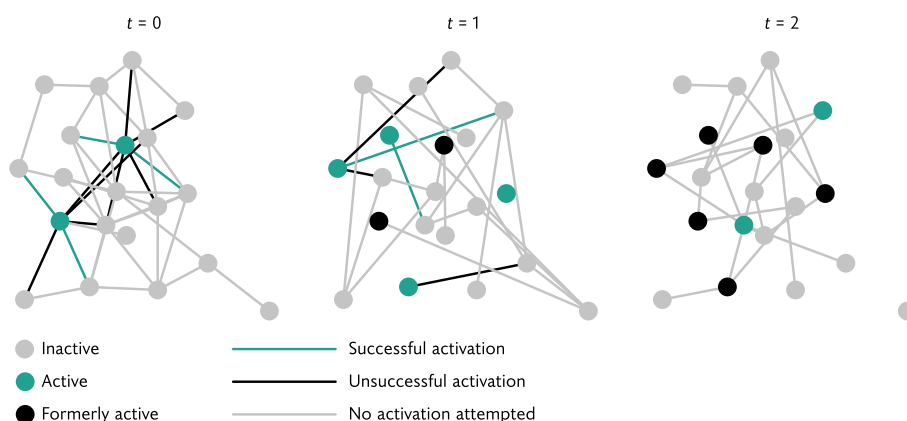


Fig. 2 Example of influence diffusion on a toy temporal network using the independent cascade (IC) model. Activated nodes (green) try to influence all of their neighbors before becoming inactive (red) in the next time step. The number of nodes currently or formerly activated corresponds to the total influence spread

the network can only be known by probing a small subset of nodes and this partially visible network is used for IM seeding. See Fig. 1 for an overview comparing these different paradigms. Real-world implementation of IM algorithms is the topic of section “[Real world implementations](#)” where we primarily focus on the problem of increasing HIV awareness amongst homeless youth. This application highlights the many challenges associated with deploying IM algorithms. Finally, we conclude in Sect. 5 with important areas for future research, including the *ex ante* setting, model misspecification, and the temporal relationship between the diffusion process and network evolution. Throughout the paper, we give special attention to the functionality of these methods on real-world problems.

Notation

We begin by defining common notations used throughout the paper. Let $G = (G_0, \dots, G_{T-1})$ be a temporally evolving network over T time stamps. Typically the graph snapshots G_t occur over evenly spaced time intervals, i.e., $t_k - t_l$ is constant for all k, l . For each t , let $G_t = (V_t, E_t)$, where V_t is the set of vertices and E_t is the set of edges. Typically, $V_t \equiv V$ and does not vary with time. Let $n = |\cup_t V_t|$ and $m = \sum_t |E_t|$ be the total number of nodes and edges in the network, respectively. Additionally, let $A_{ij}(t)$ denote the corresponding adjacency matrix for graph G_t where $A_{ij}(t) = 1$ if there is an edge from node j to node i at time t , and 0 otherwise. In an undirected network, $A_{ij} = A_{ji}$ for all i, j , while A may be asymmetric for a directed network. Let $N_i(t)$ be the set of incoming neighbors of node i at time t , i.e., $N_i(t) = \{j : A_{ij}(t) = 1\}$.

Diffusion mechanisms

In order to study IM, it is necessary to describe the diffusion of influence on a network. The most common diffusion models are the independent cascade (IC), linear threshold (LT) and susceptible-infected-recover (SIR) models. In the IC model (Kempe et al. 2003; Saito et al. 2008; Shakarian et al. 2015), influenced nodes have a single chance to activate their uninfluenced neighbors. Specifically, let p_{ij} be the

probability that node i influences nodes j . If node i is infected at time $t - 1$ and $A_{ij}(t - 1) = 1$, then node j becomes infected at time t with probability p_{ij} . From time t , node i can no longer influence its neighbors. Then the total influence spread is the number of nodes that were active at any point.

In Fig. 2, we show the influence diffusion process on a toy network using the IC model. Two nodes (green) are initially selected for the seed set and begin in the active state. These nodes attempt to activate all of their neighbors, but only some attempts are successful (green and red edges are successful and unsuccessful activations, respectively). In the next time step, the newly activated nodes now attempt to influence all of their current neighbors, while the previously activated nodes become inactive. This process continues one more time step, and the number of nodes in the active (green) or formerly active (red) state is the total influence spread for this seed set (eight nodes). Due to the stochastic nature of the process, even if the process is repeated with the same seed nodes, the total diffusion spread may differ.

The SIR model (Pastor-Satorras et al. 2015; Erkol et al. 2022) is similar to the IC model, but now each activated node has a fixed probability λ of infecting its unactivated neighbors. Moreover, a node can activate its neighbors as long as it is in the infected state. Each infected node, however, has probability μ of “recovering” and being unable to activate its neighbors. The total influence spread is the final number of nodes in the infected and recovered state. If $p_{ij} = \lambda$ for all i, j and $\mu = 1$, then the SIR and IC models are equivalent. Additionally, the SIR model reduces to the susceptible-infected (SI) model (Osawa and Murata 2015; Murata and Koga 2018) if $\mu = 0$.

In the LT model (Kempe et al. 2003; Chen et al. 2010; Pathak et al. 2010; Shakarian et al. 2015), each node is randomly assigned a threshold θ_i and each edge endowed with a weight b_{ij} . If the sum of weights for a node’s infected neighbors exceeds its threshold, then this node becomes infected, i.e., node i is activated if $\sum b_{ij} > \theta_i$ where the sum is over all infected neighbors of i .

Problem statement

With notation and diffusion mechanisms in hand, we formally define the IM problem. Let G be some dynamic network and let D be a diffusion mechanism, e.g., IC or SIR. We define $\sigma(S)$ as the expected number of influenced nodes for seed set S and for diffusion process D on graph G . Of course, the behavior of $\sigma(S)$ depends on G and D , but we suppress this dependence in the notation. For a fixed $k = |S|$, we seek the seed set S which maximizes $\sigma(S)$, i.e.,

$$S^* = \arg \max_{S \subseteq V, |S|=k} \sigma(S). \quad (1)$$

Perhaps the simplest approach to approximate (1) is evaluating $\sigma(S)$ via MC simulations and choosing the node which marginally leads to the largest gain in influence spread, as outlined in Algorithm 1. In the static setting, this greedy algorithm provably yields a result within a factor of $(1 - 1/e)$ of the global optimum (Kempe et al. 2003). MC simulations are computationally intensive, however, so many methods focus on efficiently computing the influence spread before employing a greedy algorithm.

Algorithm 1 Greedy influence maximization

Result: Seed set S
Input: Temporal network \mathcal{G} , seed size k
 $S = \emptyset$
for i **in** $1 : k$ **do**
 $v = \arg \max_{u \in V \setminus S} \{\sigma(S \cup \{u\}) - \sigma(S)\}$
 $S = S \cup \{v\}$
end

Another common paradigm for selecting seed nodes that avoids direct calculation of the influence function is based on node ranking, e.g., Michalski et al. (2011), Murata and Koga (2018), Michalski et al. (2020). Nodes are ranked based on some measure of importance like degree or centrality, and the k nodes with the largest value are chosen for the seed set. While much faster than greedy algorithms, these approaches yield no theoretical guarantees and may choose nodes that “overlap” their influence. For example, if two nodes have high degrees but share many common neighbors, then seeding both nodes may not be optimal as their influence will spread to the same nodes.

Reverse Reachable (RR) sketches also choose seed nodes without direct computation of $\sigma(S)$ Cohen et al. (2014), Kim et al. (2017), Guo et al. (2020). For a given time t and for each edge (i, j) , we randomly draw $Z_{ij} \sim \text{Bernoulli}(p_{ij})$ where p_{ij} is the probability that node i influences node j , and keep the subgraph with $Z_{ij} = 1$. These edges are sometimes referred to as the “live” or “active” edges. Once this subgraph is constructed, the source and destination of each (directed) edge are reversed before randomly selecting a node. Finally, a breadth-first search is conducted from this randomly selected node and all nodes reached by this search are kept for this particular RR-sketch. Essentially, the nodes in this set are those that can influence the selected node through the diffusion process. This process is repeated a large number of times to yield a set of RR-sketches.

Single seeding

The classical IM problem is where a practitioner selects a set of seed nodes at time $t = 0$ in order to maximize the influence spread at time $t = T$. Most solutions either estimate the influence spread via probabilities, or use some heuristic to rank nodes by influence. For the majority of these methods, the complete temporal evolution of the network is assumed to be known, but some relax this assumption. We also discuss several works which do not present novel algorithms, but rather analyze existing methods and/or diffusion processes.

Algorithms

First, we discuss algorithms which solve the single seeding temporal IM problem.

Greedy

The greedy algorithm of Kempe et al. (2003) naturally extends to the temporal setting: nodes with the largest marginal gain in influence spread are added to the seed set incrementally as in Algorithm 1. The only difference from Kempe et al. (2003) is that the expected influence spread is computed on a temporally evolving network. This

method is considered the “gold standard” of IM algorithms and can easily be adapted to any diffusion model. On the other hand, this algorithm suffers from a high computational cost due to the repeated MC simulations required for computing the influence spread.

Probability of influence spread

Since the costly step of the greedy algorithm is computing the influence spread, several heuristics exist which use the probability of a node’s activation in order to approximate $\sigma(S)$. In Aggarwal et al. (2012), $\pi_i(t)$ is the probability that node i is activated at time t . Assuming the network is a tree, if $p_{ji}(t)$ is the probability that node j activates node i at time t , then the probability that node i is activated at time $t + 1$ is

$$\pi_i(t + 1) = \pi_i(t) + (1 - \pi_i(t)) \times \left(1 - \prod_{j \in N_i(t)} (1 - p_{ji}(t)) \right) \tag{2}$$

where $N_i(t)$ is the set of incoming neighbors of node i at time t . We explicitly write $N_i(t)$ as a function of t to stress that the neighboring set of each node is time-dependent and thus encodes the temporal variation in the network. The first term in (2) is the probability that the node was already activated during the previous time step while the second term is the probability that it was not previously activated but becomes so in the current step. The authors initialize $\pi_i(1) = 1$ if $i \in S$ and 0 otherwise. For each i , $\pi_i(t)$ iteratively updates via (2) and $\sum_i \pi_i(T)$ estimates $\sigma(S)$. Aggarwal et al. use this procedure and a greedy algorithm to choose the seed set. The authors also assume that $p_{ij}(t) := p_{ij} = p_{ij}(\Delta_{ij}^{(T)})$ is an increasing function of the total amount of time that an edge between nodes i and j exists in the network, $\Delta_{ij}^{(T)}$. This means the edge transmission probabilities do not vary temporally, but only depend on $\Delta_{ij}^{(T)}$. For example, if $A_{ij}(0) = A_{ij}(1) = 1$, $A_{ij}(t) = 0$ for all $t > 1$, then $\Delta_{ij}^{(T)} = 2$, and $A_{kl}(0) = 1$, $A_{kl}(t) = 0$ for all $t > 0$, then $\Delta_{kl}^{(T)} = 1$, so $p_{ij}(\Delta_{ij}^{(T)}) > p_{kl}(\Delta_{kl}^{(T)})$. Additionally, this method eschews the standard, equally-spaced graph snapshots in favor of times corresponding to structural changes based on the number of edge updates. The approach also can find the most likely seed nodes for a given diffusion pattern.

Osawa and Murata (2015) take an analogous approach to Aggarwal et al. but use the SI model for diffusion. In fact, this method is equivalent to Aggarwal et al. (2012) if $p_{ij}(t) = \lambda$ for all i, j and times t . Osawa and Murata show this approach slightly overestimates the true influence spread and prove that the associated greedy algorithm’s computational complexity is $O(nmk)$. In simulations, this method outperforms broadcast (Grindrod et al. 2011) and closeness centrality (Holme and Saramäki 2012) heuristics. It also yields comparable performance to a standard greedy algorithm but is two orders of magnitude faster. Additionally, Osawa and Murata show that centrality heuristics perform worse on networks with strong community structure due to nodes’ overlapping influence.

Erkol et al. (2020) extend this paradigm to the SIR model. If $\pi_i(t)$ is defined as above and $\rho_i(t)$ is the probability that node i is in state R at time t (such that $1 - \pi_i(t) - \rho_i(t)$ is the probability of being in state S), then the SIR dynamics are defined by

$$\pi_i(t) = (1 - \mu)\pi_i(t - 1) + (1 - \pi_i(t - 1) - \rho_i(t - 1)) \times \left(1 - \prod_{j \in N_i(t)} (1 - \lambda\pi_j(t - 1)) \right) \tag{3}$$

$$\rho_i(t) = \rho_i(t - 1) + \mu\pi_i(t - 1). \tag{4}$$

A greedy algorithm chooses the seed nodes based on the influence spread estimate $\sum_i \{\pi_i(T) + \rho_i(T)\}$. Erkol et al. study the performance of the method when the network is noisy or incomplete, i.e., the temporal snapshots are randomly re-ordered, only the first snapshot is available, and the network is aggregated into a single snapshot. The authors find that the order of snapshots is crucial and ignorance about G_0 causes the algorithm to suffer. Indeed, in many cases, knowing only G_0 is sufficient for large influence spread, while the aggregation approach consistently performs poorly. Erkol et al. also show that if the recovery probability, μ , is large, then central nodes in the first few layers are the best influence spreaders, whereas for small μ , nodes must be central in many layers to make for optimal seed nodes.

Node ranking heuristics

Rather than estimate the influence spread, the following methods rank nodes by influence and select the top k as seed nodes. The earliest approach comes from Michalski et al. (2014). The authors adopt the LT model and assume the first $T/2$ snapshots are available to select seed nodes, but the diffusion process occurs on $G_{T/2}, \dots, G_{T-1}$. Some static measure of node importance $m_i(t)$ is computed for all nodes i and snapshots G_t . The values across t are combined by down-weighting older values to yield a single metric θ_i for each node, i.e.,

$$\theta_i = \sum_{t=0}^{T/2-1} f(m_i(t), t) \tag{5}$$

where $f(x, t)$ is an increasing function in t . The k nodes with largest θ_i for a given $f(\cdot)$ are chosen as seed nodes. Michalski et al. find that the following combinations of metrics $m_i(t)$ and forgetting mechanisms $f(\cdot)$ yield the largest influence spread²: out-degree and in-degree with exponential forgetting ($f_{exp}(x, t) = e^{-t}x$) total degree and logarithmic forgetting ($f_{log}(x, t) = \log_{T/2-1-t+1}(x)$), betweenness centrality and hyperbolic forgetting ($f_{hyp}(x, t) = (T/2 - 1 - t - 1)^{-1}x$); and closeness centrality with power forgetting ($f_{pow}(x, t) = x^t$). The authors also vary the number of aggregated snapshots used to compute the metrics and find that the finest granularity performs the best. Indeed, treating the network as static by aggregating $G_0, \dots, G_{T/2-1}$ into a single graph yields the lowest influence spread on $G_{T/2}, \dots, G_{T-1}$, thus demonstrating the importance of accounting for the temporal variation in the network.

Another node ranking heuristic comes from Murata and Koga (2018). Using the SI model, the authors extend several static measures of importance to the temporal

² The exponential forgetting mechanism is as reported in the original paper, but there appears to be some error since this function clearly *decreases* with t .

settings. In particular, the dynamic degree discount algorithm extends (Chen et al. 2009). First, the node with largest dynamic degree $D_T(v)$ is added to the seed set, where

$$D_T(v) = \sum_{t=2}^T \frac{|N_v(t-1) \setminus N_v(t)|}{|N_v(t-1) \cup N_v(t)|} |N_v(t)| \quad (6)$$

and $N_v(t)$ is the neighbors of node v at time t . Once a node is selected, the value of the dynamic degree for its neighboring nodes is decreased and the process repeats until k nodes are selected. The authors show that the complexity of this method is $O(k \log n + m + mT/n)$ but that it only is valid for the SI model. Murata and Koga also propose Dynamic CI as an extension of (Morone and Makse 2015) based on optimal percolation and Dynamic RIS as an extension of Borgs et al. (2014), Tang et al. (2014) based on RR sketches. In simulations, all methods perform comparably to Osawa and Murata (2015) but are significantly faster. The authors also show that when λ is large, choosing the optimal seed nodes is less important as many seed sets yield comparable influence spread.

Recently, Michalski et al. (2020) propose another node-ranking heuristic. The authors postulate that, for the IC model, nodes with large variability in their neighbors should be chosen to maximize spread. They quantify neighborhood variability with an entropy measure that rewards nodes for changing their neighbors in subsequent graph snapshots and the k nodes with the largest value are chosen for the seed set. The measure is computed on the first $T/2$ snapshots while the influence is calculated on the second half of the graph's evolution, similar to Michalski et al. (2014). The authors note, however, that this metric may not make sense for the LT model which requires the number of activated neighbors to "build up" for a node to become infected. As the method depends on the neighborhood set of each node, its complexity is $O(m)$.

To summarize the methods in the previous two subsections, we include Table 1 which compares them across several metrics.

Analysis

We turn our attention to methods that do not propose a novel IM algorithm, but rather analyze the existing algorithms and/or diffusion mechanisms. In order to better model information propagation, Hao et al. (2011) propose two novel diffusion models where a node's propensity of activation depends on the number of past attempts to activate it. In the time-dependent comprehensive cascade model, an active node still only has a single chance to activate its neighbors, but the probability of being infected can either increase, decrease or be unaffected by the number of previous attempts on that node. The authors also propose a dynamic LT model where the node's activation threshold depends on the number of previous activation attempts. Hao et al. proposes a time series-based approach to empirically determine the effect of past activation attempts on infection probabilities.

Gayraud et al. (2015) study the behavior of the influence spread function under several novel diffusion models while also allowing seed nodes to be activated at different times. Let $f : 2^V \rightarrow \mathbb{R}$ be a set function. If $S \subseteq V$, then f is

Table 1 Comparison of different algorithms for single seeding temporal influence maximization

Method	Paradigm	Model	Complexity
Greedy	Greedy	Any	$O(nmkR)$
Aggarwal et al. (2012)	Approx.	ICt	$O(nmk)$
Osawa and Murata (2015)	Approx.	SI	$O(nmk)$
Erkol et al. (2020)	Approx.	SIR	$O(nmk)$
InExp Michalski et al. (2014)	Node rank	LT	$O(m)$
OutExp Michalski et al. (2014)	Node rank	LT	$O(m)$
TotalLog Michalski et al. (2014)	Node rank	LT	$O(m)$
BetHyp Michalski et al. (2014)	Node rank	LT	$O(nm + n^2T)$
CloPow Michalski et al. (2014)	Node rank	LT	$O(nm + n^2T)$
Dynamic Degree Murata and Koga (2018)	Node rank	SI	$O(k \log n + m + mT/n)$
Dynamic CI Murata and Koga (2018)	Node rank	SI	$O(n \log n + mT/n)$
Dynamic RIS Murata and Koga (2018)	Node rank	SI	$O(\theta dkl^2(m+n) \log^2 n/\epsilon^3)$
Entropy Michalski et al. (2020)	Node rank	IC	$O(m)$

Paradigm: *Approx.*: estimates the probability that a node is activated. *Node rank*: uses a node ranking heuristic. Model: Diffusion model. Note that ICt refers to Aggarwal et al. (2012)'s model which does not neatly line up into any of the standard models but it similar to an IC model with time component. Complexity: number of flops to implement the algorithm. R is number of MC simulations; θ, d, l are tuning parameters in the Dynamic RIS algorithm

monotone if $f(S \cup \{v\}) - f(S) \geq 0$ for all $v \in V \setminus S$ and $S \subseteq V$. It is *submodular* if $f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)$ for $A \subseteq B$ and $v \in V \setminus B$. In other words, the monotone property implies that adding a node never decreases the influence spread and submodularity means that there is diminishing returns for adding more nodes. Additionally, the authors define a seeding strategy to be *timing insensitive* if all nodes should be activated at time $t = 0$ and *timing sensitive* otherwise. In the transient evolving IC model (tEIC), infected nodes at time $t - 1$ have one chance to infect their neighbors at time t . The authors prove that this diffusion mechanism is neither monotone nor submodular. In contrast to the tEIC model, the persistent EIC model assumes that a node tries to activate its neighbors the first time that the two nodes have a link. If the activation probabilities are constant in time, Gayraud et al. prove that this model is monotone, submodular and timing insensitive. If the activation probabilities dynamically vary, then the influence function is neither monotone nor submodular and is timing-sensitive. The authors propose similar extensions to the LT model. The transient ELT model only considers weights from active neighbors at the current snapshot whereas the persistent ELT model sums all weights from neighbors activated during any previous time. These models are monotone, not submodular and timing insensitive, and monotone, submodular and timing insensitive, respectively. The key contribution of this paper is that if a model is timing-sensitive, the seed nodes should be activated throughout the diffusion process, as opposed to all at $t = 0$. The authors also show that choosing seed nodes based on aggregating all graph snapshots does not perform well for any model.

The submodularity of the influence function is also studied in Erkol et al. (2022), this time under the SIR model. The authors show that if $\mu = 0$ (SI model), the influence function is submodular, but loses this property when $\mu > 0$. Effectively, the violations

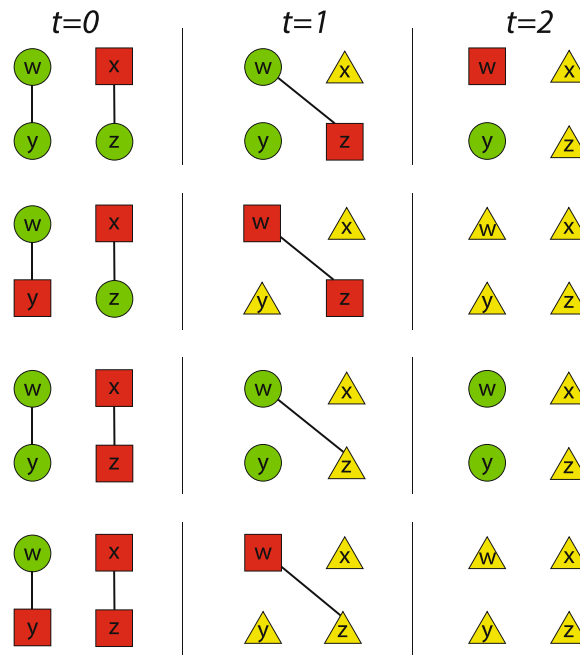


Fig. 3 Toy example showing the violation of the sub-modularity property in the temporal SIR model. Green circles are susceptible nodes, red squares are infected nodes, and yellow triangles are recovered nodes. Each row represents the diffusion process for a particular seed set with $\lambda = \mu = 1$. In the first row, $S_1 = \{x\}$ and the total influence spread is three nodes. In row three, $S_2 = \{x, z\}$, but the total influence spread is only two nodes. Thus, $S_1 \subset S_2$ while $\sigma(S_1) > \sigma(S_2)$. Reproduced from Erkol et al. (2022) with permission of the author

come from nodes in state R “blocking” paths to nodes in state S, as demonstrated by a toy example in Fig. 3 (reproduced with permission of the author).

A relaxation of the submodularity property, γ -weakly submodular, is also not achieved in the SIR model. A function f is γ -weakly submodular if for $A \cap B = \emptyset$ and $0 < \gamma \leq 1$,

$$\sum_{v \in B} f(A \cup \{v\}) \geq \min \left\{ \gamma f(A \cup B), \frac{1}{\gamma} f(A \cup B) \right\}. \tag{7}$$

The authors then empirically check the number of violations of the submodular criteria in real networks. They find that if nodes are randomly selected, the criteria is frequently violated. If nodes are selected based on a greedy algorithm, however, the submodularity property is rarely violated. Thus, the influence function is effectively submodular. Now, since the influence function is not submodular, there is no theoretical guarantee that the greedy algorithm adequately approximates the optimal solution. In spite of this, compared with a brute-force algorithm on real-world networks, the greedy algorithm still yields results within 97% of the optimal solution.

Lastly, although not pertaining explicitly to IM, we briefly discuss Albano et al. (2013). This work studies the relationship between graph topological evolution and diffusion processes by analyzing which part of the diffusion is owed to the diffusion mechanism, and which to graph dynamics. The authors consider two timing mechanisms: *extrinsic* time based on seconds between interactions, and *intrinsic* time based on changes or transitions in the network. While researchers typically use extrinsic time, the authors argue that intrinsic time may be more sensible in many cases. Using the SI model and

intrinsic time, the observed diffusion is governed more by the diffusion mechanism than the evolution of the network. Using extrinsic time, conversely, the topological changes in the network greatly affect the diffusion. Thus, the diffusion process is highly dependent on the timing method.

Discussion

In the previous subsections, we presented the leading methods for choosing a single seed set in temporal IM. Each method either estimates the influence spread, or ranks nodes based on a heuristic. Aggarwal et al. (2012), Osawa and Murata (2015) and Erkol et al. (2020) proposed analogous approaches with the only difference being in the diffusion model. These methods maintain many of the desired properties of the greedy algorithm, but are computationally less intensive. The node ranking metrics of Michalski et al. (2014), Murata and Koga (2018) Michalski et al. (2020) are even faster since they avoid the costly influence spread calculation.

In practice, if a network is small enough, the greedy algorithm should always be preferred; it comes with theoretical guarantees and empirically yields the largest seed set. Indeed, the majority of real-world applications discussed in Sect. 4 employ a greedy approach. When there is not enough computing time or resources for the greedy algorithm, the probability-based methods are the second best choice. Osawa and Murata as well as Erkol et al. show that these tend to yield seed sets which lead to larger average outbreak sizes than those based on node ranking. Indeed, the influence spread of Erkol et al.'s algorithm comes within 95% of that of the greedy algorithm under different settings of the SIR model. On the contrary, closeness centrality-based ranking methods might only have influence spreads less than one third that of greedy's (Osawa and Murata 2015). Node ranking methods should be reserved for extremely large networks where the previous two frameworks are infeasible.

Knowledge of the network structure and diffusion process can also inform algorithm selection. For networks without a community structure, Osawa and Murata show methods which approximate the influence spread function do not outperform those based on node ranking heuristics. Similarly, when there is a high probability of nodes recovering in the SIR model, then the most central nodes in the first few snapshots of the graph are the best to include in the seed set (Erkol et al. 2020). Finally, if the infection probability is large, then Murata and Koga find the algorithm choice is of little consequence as many seed sets similar influence spreads.

A key challenge in implementing these methods on real-world problems is the requirement that the entire topology of the network be known. Save (Michalski et al. 2014, 2020), each method assumes that the evolution of the network G_0, \dots, G_{T-1} is known at time $t = 0$ when the seed nodes are selected. Of course, in practice, it is unreasonable for a practitioner to know the future topology of the network, so it is not obvious how to apply these methods in this case. To address this issue, Yanchenko et al. (2023) propose a link prediction approach for *ex ante* temporal IM. Using the SI model, the authors use the first p snapshots to train a link prediction algorithm and then predict the network topology for G_p, \dots, G_{T-1} . An existing temporal IM algorithm is applied to these predicted networks to choose the seed sets. In many cases, finding seed nodes on a simple aggregation of G_0, \dots, G_{p-1} performs as well as the more complicated link prediction

methods. This finding is at odds with Michalski et al. (2014) and Erkol et al. Erkol et al. (2020) who showed poor performance of IM algorithms on aggregated networks. These papers, however, assumed different diffusion mechanisms, so it is possible that aggregating only works well for the SI model. Another practical consideration for applied researchers is the size of the network. If working with a relatively small social network, then a greedy algorithm is reasonable, whereas a node ranking heuristic is mandatory for large online social networks with millions of nodes. Finally, the diffusion mechanism must be carefully chosen based on the application's domain, as certain methods are only applicable to specific mechanisms.

Multiple seedings

In the previous section, we considered IM algorithms for temporal networks where all seed nodes are activated at time $t = 0$. Now we discuss methods where nodes are seeded at different points throughout the evolution of the network, or where the seed set is updated at each time step.

Sequential seedings

Related to the single seeding problem, consider a single seed set S , but instead of activating all nodes at $t = 0$, nodes activate *sequentially* as the network evolves. This problem involves not only choosing *which* nodes to include in the seed set, but also *when* to activate them.

Michalski et al. (2020) focus on the seed activation step of this problem. The authors consider a variant of the IC model where a single node is activated and the diffusion occurs until no more activations are possible. Then the next node is activated and the process continues. In this setting, Michalski et al. use a simple seed selection method based on degrees. First, the node with the largest degree is activated. Once the diffusion process finishes, the uninfected node with largest degree is activated and the process continues until k nodes have been seeded. This method is compared with activating the k nodes with largest degree at time $t = 0$. When t is small, activating all nodes at once leads to a larger influence spread, but as t increases, the sequential seeding strategy outperforms the single seeding, as shown in Fig. 4 (reproduced with permission of the author).

Tong et al. (2016) consider another variation of the sequential seeding problem where seed set nodes are unsuccessfully activated with some probability. They propose a greedy algorithm which maximizes the marginal gain in influence spread given the current diffusion. Towards this end, the authors derive a closed-form expression for the expected number of influenced nodes by constructing an auxiliary graph with extra nodes and edges based on possible seed sets and propagation probabilities. The greedy algorithm is shown to yield results within $(1 - 1/e)$ of the optimal influence spread while the computational burden is mitigated with the Lazy-forward technique (Leskovec et al. 2007). Additionally, Tong et al. prove that the strategy outlined in Michalski et al. (2020) of seeding nodes one at a time and waiting for the diffusion process to finish before activating the next node is the optimal seeding strategy for any temporal graph.

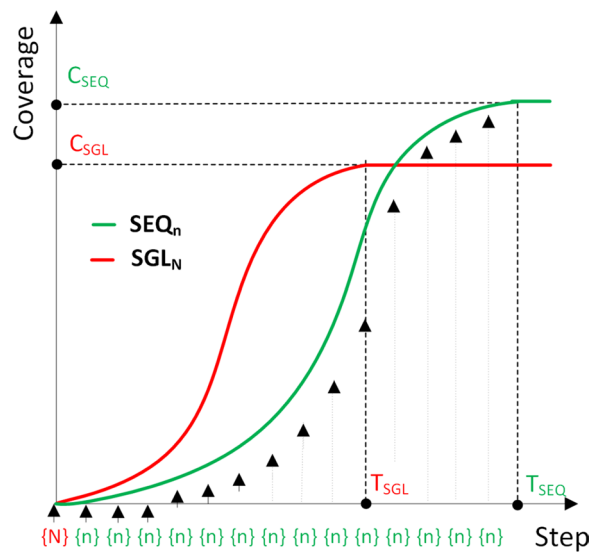


Fig. 4 Comparison of the influence spread for the sequential seeding strategy with that of single seeding. The red curve activates all nodes at $t = 0$ while the green curve activates nodes one at a time. Single seeding leads to greater influence spread at first, but the sequential seeding strategy ultimately leads to more spreading. Reproduced from Michalski et al. (2020) with permission of the author

Maintenance seeding

In a highly dynamic network, the optimal seed set may change with time. For example, in a long-term marketing campaign on X , the active users and followers change over the duration of the campaign. Thus, it is necessary to *maintain* or *update* the seed set S_t such that it provides maximum influence spread on G_t for all t . This problem is known as *maintenance seeding*. Maintenance seeding is markedly different from static seeding as now k nodes are activated at each time step t in order to maximize the diffusion on G_t . Thus, this process is analogous to a sequence of static IM problems.

Chen et al. (2015), Song et al. (2016) first study this problem under the name “influential node tracking.” The authors assume that the topology of the network is known at the next time step G_{t+1} and use the IC model for diffusion. Using the seed set from the current snapshot S_t , Chen et al. employ an interchange heuristic (Nemhauser et al. 1978) to efficiently update the seed set and prove that the solution is guaranteed to be within $1/2$ of the optimal spread. Effectively, this method swaps one node in S_t with one node in $V \setminus S_t$ to maximize the marginal gain in influence spread. Since evaluating the marginal gain for every node in $V \setminus S_t$ is expensive, the authors only consider nodes with the largest marginal gain upper bound. If the upper bound for node $u \in V \setminus S$ is smaller than the marginal gain of another node v , then evaluating the influence of node u is unnecessary as its inclusion cannot improve the total influence spread. The proposed algorithm has $O(kn)$ complexity.

Ohsaka et al. (2016) consider a similar problem for large online networks where nodes and edges are added or removed at each time step. Using the IC model, the authors propose a sketching method akin to RR sets and an efficient data structure to build and store these sketches. A greedy algorithm is then implemented to choose the seed sets. Specifically, the node which is present in the most sketches is chosen as a seed node. Then all sketches which contain that node are removed from consideration, and the node which

occurs in the most remaining sketches is chosen for the seed set. This process continues until k nodes are chosen. In addition to the novel data structure, this work proposes heuristics that lead to efficient updates of the sketches at each evolution of the graph, instead of recomputing them from scratch. These heuristics come with theoretical guarantees and lead to algorithmic speed-ups.

There are several other methods that address this problem. Wu et al. (2019) recast it as a bandit problem but tackle it in a similar manner to Ohsaka et al. by using RR sketches. Wang et al. (2017) find the optimal seed nodes at $t = 0$ and incrementally update them based on investigating parts of the graph which changed significantly between snapshots. Wang et al. (2017) selects seed nodes based on a sliding window scheme, and Chandran and Viswanatham (2022) uses a node's number of triangles to estimate its influence. Min et al. (2020) consider a special case by accounting for user attributes in an online social network, including preferred topics of engagement. The authors also account for certain time periods where users may be inactive and allow for a different diffusion model based on the topic.

Up to this point, each method assumes knowledge of the future topology of the network. Singh and Kailasam (2021) relax this assumption by predicting the graph structure one time step in the future using a conditionally temporal restricted Boltzmann machine (Li et al. 2014) and then finding the seed nodes on the predicted graph. The authors use an interchange (Nemhauser et al. 1978) heuristic to update the seed set and ideas from (Song et al. 2016) to improve efficiency.

Rather than propose a new maintenance seeding algorithm, Peng (2021) studies the amortized running time, i.e., the amount of time it takes to update the seed nodes at each time step. Even though the current algorithms efficiently update the seed set in $O(n)$ time for each t , the author argues that this is still too slow for large networks. Peng then considers two different graph evolution paradigms, both under either the IC or LT model. First is an *incremental* model where a network may only add new nodes and edges. Under this model, Peng shows that an $(1 - 1/e - \epsilon)$ approximation of the optimal solution is possible with probability $1 - \delta$ for amortized running time $O(k\epsilon^{-3} \log^3(n/\delta))$, much faster than $O(n)$. Under a fully dynamic model, where nodes and edges can be added and deleted, the author proves that a $2^{(-\log n)^{1-o(1)}}$ approximation is impossible without $n^{1-o(1)}$ amortized run time. Thus, there is no possibility of improving the $O(n)$ run time.

Node probing

While previous methods assumed complete knowledge of the future network topology, the *node probing* problem assumes that the future graph snapshots are unknown but can be partially observed by probing the neighborhoods of certain nodes. Here, probing a node means observing its edges. Assuming G_0 to be known, the goal is to carefully select which nodes to probe in order to have the most information on the topology of the network in order to effectively implement an IM algorithm. This problem may arise in large online social networks where it is infeasible to observe the activities of all users at every time step. Another relevant application is modeling the social connections within a hard-to-reach population, e.g., homeless youth, as there is no straightforward way to observe all the people (nodes) in this network, yet alone the friendships (edges).

This problem was originally formulated by Zhuang et al. (2013). For each t , the researcher probes b nodes and observes changes in their neighborhoods. Once the nodes are probed, an IM algorithm is implemented on the (incomplete) visible network. Thus, the goal is to find the ideal probing strategy. The authors propose probing nodes that yield the maximum possible change to the solution of the IM problem. Since the authors use the degree discount algorithm (Chen et al. 2009), this reduces to finding the nodes with greatest change in their degree. Specifically, let $\beta(v)$ be the maximum difference in the influence spread of optimal seeds chosen before and after probing node v . Moreover, let S be the optimal seed nodes at time $t - 1$ and let S_0 be the k nodes with the largest in-degree on the most up-to-date graph snapshot. Let $t - c_v$ be the last time stamp at which node v was probed. For $\epsilon > 0$, if $z_v = \sqrt{-2c_v \log \epsilon}$, $\beta(v)$ is derived as:

$$\beta(v) = \begin{cases} \max\{0, \max_{u \notin S} \hat{d}_{in}(u) - \hat{d}_{in}(v) + z_v\}, & v \in S_0 \\ \max\{0, \hat{d}_{in}(v) - \min_{u \in S} \hat{d}_{in}(u) + z_v\}, & v \notin S_0 \end{cases} \quad (8)$$

where $\hat{d}_{in}(v)$ is the in-degree of node v based on the most recently probed network. Then node $v^* = \arg \max_{v \in V} \beta(v)$ is probed and the network topology is updated. Once b nodes have been probed, the degree discount algorithm is applied to determine the optimal seed nodes for influence spread.

Han et al. (2017) study the same problem but focus on communities with high variation as opposed to nodes. The authors postulate that the total in-degree for a community should be relatively stable with time, so if this changes greatly, there must have been a significant change in this community and it is worth probing. The authors use the community detection algorithm of Zhou et al. (2009), and once the community with high variability has been identified, they employ a probing algorithm similar to that of Zhuang et al. (2013).

Discussion

We close this section by highlighting important considerations for practical implementation of these methods. In the sequential seeding setting, it is important for researchers to consider how long they can allow the diffusion to take place since static seeding is preferable for small T and sequential for large T . Michalski et al. (2020) also emphasize that the sequential strategy is better suited for independently activated models, e.g., IC and SI, rather than threshold-based models, e.g., LT, so the diffusion model is another important consideration. It would also be interesting to compare static and sequential seeding for more complicated IM algorithms. It is well-known that seeding the top k degree nodes is a relatively poor IM algorithm, so it is unclear whether sequential seeding would perform so much better when combined with different IM algorithms.

For maintenance seeding, we observe that the seeding budget is effectively kT rather than k , since k nodes are activated at T different time steps. If T is large, then it may be prohibitive to keep activating k nodes each round. Additionally, this setting implicitly assumes that nodes can be reinfected at successive snapshots, i.e., $S_t \cap S_{t+1} \neq \emptyset$. This may be reasonable in epidemiological settings, for example, where a person can be reinfected by a disease. For marketing campaigns, on the other hand, it is unlikely that a user targeted with an ad in multiple time steps can be expected to have significant diffusion

in each case. Thus, the number of times that a user has been infected and this effect on the diffusion mechanism should be considered carefully. Moreover, save the IC model, if a node is infected at time t , then it could continue to attempt to infect its neighbors at $t + 1, t + 2, \dots$. The frameworks presented above, however, assume that unless nodes are in the new seed set S_{t+1} , they are unable to exert influence. Next, save (Singh and Kailasam 2021), each maintenance seeding method assumes that the future network topology is known, which is generally untrue in practice. In particular, sequential seeding strategies when the graph snapshots are unknown is an important and practically relevant open problem. Finally, Yang et al. (2017) argue that identifying influential nodes is a separate task from influence maximization. For example, if a new user joins X , they may want to follow the most influential users. Identifying these users is different from trying to maximize the spread of a product or idea on X 's network.

The node probing problem is a promising step toward practically relevant IM algorithms. Indeed, assuming that the network structure is unknown, except through probing, is much more realistic than the methods which assume complete topological information. These methods, however, treat the problem as a sequence of static IM tasks since the seed nodes are computed fresh at each time step. An interesting advance would be to leverage the previous seed set in computing the new seeds.

Real world implementations

A key focus of this review is understanding if existing methods are prepared to handle IM tasks “in the field.” By real-world implementation, we mean that researchers intervene to initially influence nodes in the network and then empirically calculate or estimate the information spread at the end of the experiment. This contrasts with most IM studies which may use a real network, but then perform simulations on these networks to both influence nodes and compute the final spread.

To date, the literature on IM in real-world settings is scant. In this section, we highlight the existing studies and discuss some of the associated challenges. While these works assume that the network is static, the majority employ a sequential seeding strategy which is why we include it in our discussion of temporal IM. To our knowledge, there are no existing papers explicitly implementing IM algorithms on dynamic networks.

The most notable examples of applied IM comes from a series of papers by Yadav et al. (2016), Yadav et al. (2017), Yadav et al. (2018), Wilder et al. (2017), Wilder et al. (2018). In these works, the goal is to maximize HIV awareness among homeless youth in large urban areas. This is a classic IM setting as homeless shelters can only train a small number of youth on HIV prevention, but hope that participants pass this information along to their friends to maximize awareness. The general problem setup is as follows. First, the social network of homeless youth is partially constructed. Then the homeless shelter chooses k youth to participate in an intervention on HIV prevention. During the training, the youth reveal all of their one-hop friendships. The information is then given time to diffuse on the network (but this spread is unknown) before inviting k more youth for training. This process continues for T training rounds.

There are several key challenges to deploying IM algorithms in this setting. First, the complete social network of the homeless youth population is unknown, both in terms of nodes (youth) and links (friendships). Moreover, new information on the network

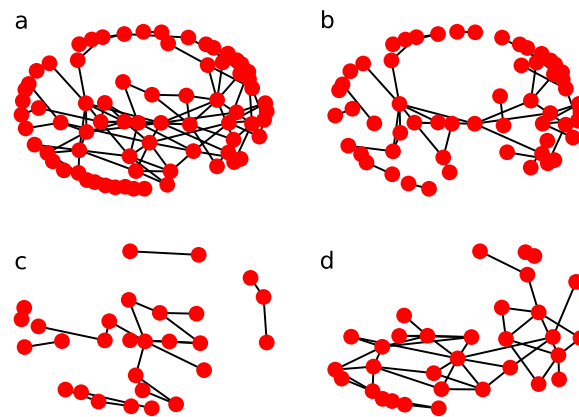


Fig. 5 Homeless youth social network constructed using different methods. **a** All methods combined. **b** Self-reported edges. **c** Field observations. **d** Staff observations. Reproduced from Wilder et al. (2018) with permission of the author

structure is collected during the experiment as youth are trained and their friendship circle is elucidated. Second, youth may refuse and/or be unable to attend the training, meaning that seed nodes have a certain probability of remaining inactive. Lastly, quantifying the information spread on the network is highly non-trivial. Thus, this problem combines node probing, as the network structure is partially unknown before selecting a node to learn their social circle, and sequential seeding, where the nodes are activated over time. It differs from the standard node probing problem, however, in that nodes are chosen to optimize influence spread, rather than maximize topological information about the network; it differs from sequential seeding in that the influence spread is unknown when selecting the next seed nodes.

The first attempt to address these challenges comes from Yadav et al. (2016). Assuming the SI model³ for diffusion, the authors construct the social network using Facebook friendships while inferring missing links using link prediction techniques (Kim and Leskovec 2011). They prove that the task of choosing k seed nodes at each of the T time steps is NP-hard and that it is impossible to achieve a $n^{-1+\epsilon}$ approximation of the optimal solution with an uncertain network. The problem is then recast as a Partially Observable Markov Decision Process (POMDP). By simulating the diffusion process, the nodes with the largest expected reward (influence spread) are selected for the seed set. In order for the method to handle real-world network sizes, the authors propose a divide-and-conquer approach. Their proposed method is one hundred times faster than existing methods while also yielding greater influence spread. In Yadav et al. (2016), the authors generalize the model by allowing for greater uncertainty in the influence and edge probabilities.

In Wilder et al. (2018), the authors focus on several practical considerations for this problem. First, the algorithm accounts for a non-zero probability that a seed set node remains inactive, i.e., the youth does not attend the training. The authors also address the network construction step by proposing a network sampling approach based on the *friendship paradox* (Feld 1991). This paradox says that, on average, a random node's

³ In the paper, Yadav et al. state that they use the IC model, but in terms of the notation of this paper, it falls under the SI classification.

neighbor has more friends than the original node. Thus, with a sampling budget of M nodes, they first randomly sample $M/2$ nodes and then randomly sample one neighbor per node. This approach increases the likelihood that central (e.g., influential) nodes are sampled. Figure 5 shows the homeless youth social network constructed using different methods (reproduced with permission of the author). These four networks highlight the challenges of constructing the network for a hard-to-reach population as the topology varies greatly depending on the collection method (self-report, field observations and homeless shelter staff observations). Next, the authors assume that the influence propagation probability is unknown but modeled to maximize the worst-case ratio between the true spread and the estimated spread. Finally, the authors propose a greedy algorithm to select the optimal seed nodes and prove that it is guaranteed to output a solution within a factor of $(e - 1)/(2e - 1)$ of the optimal. In a real-world pilot study, by sampling only 15% of the nodes, the proposed method achieved comparable spread compared to that if the entire network was known.

These methods are applied to the real-world task in Yadav et al. (2017), Yadav et al. (2018). Some of the key questions considered are: Do the activated nodes actually pass their information along to others? Do the activated nodes give meaningful information about the social network? Can these algorithms do a better job of selecting seed nodes than an expert (social worker) can? To answer these questions, the authors implement the methods from Yadav et al. (2016) and Wilder et al. (2017). They also consider a baseline IM algorithm based on largest degrees. For each method, the authors recruit study participants, construct the network, activate nodes (via training) and conduct follow-ups to evaluate the final influence spread. The proposed methods in Yadav et al. (2016) and Wilder et al. (2017) yielded much larger influence spreads than the degree-based method while also leading to a change in participants' behavior, i.e., increase in participants testing for HIV.

Lastly, we discuss an application of IM to an online setting. Huang et al. (2022) consider a "closed" social network where posts are only shared with certain people rather than all of the users' connections. As a slight variation to the standard IM problem, the authors find the friends with which the user should share their information to maximize spread. In other words, the goal is to maximize the influence spread where users can only share the information with a limited subset of edges (neighbors). Indeed, this may be a more realistic diffusion mechanism for social networks as it is unlikely that someone would give equal effort to share information with each of their friends; rather, he/she would likely target a few specific people. The authors apply this to an online multiplayer game where each user is recommended friends to interact with, e.g., send gifts, game invitations, etc. The proposed method is compared with randomly selecting friends and yields a 5% increase in click-through rate.

Considerations and future directions

We concluded by sharing thoughts on the challenges associated with temporal IM as well as some of the important areas for future research.

Real-world implementations

In Sect. 4, we saw the litany of challenges facing a researcher trying to implement IM algorithms on real-world problems. We list a handful of questions that he/she must consider in applying these methods: What is the information diffusion mechanism? Can nodes be sequentially updated, or are they all activated at the start? Will seed nodes be activated with certainty? How long does the diffusion process continue? On what time scale is the network evolving? How long does it take to influence a node? Are the network dynamics changing rapidly? Does the future topology of the network need to be predicted? Is the network updated in an online setting or with standard snapshots? Is the true influence spread known? We look forward to many more IM implementations in real-world applications.

Single seeding methods

In Sect. 2, we discussed several methods for the single seeding temporal IM problem. There were only five papers, however, and Aggarwal et al. (2012), Osawa and Murata (2015) and Erkol et al. (2020) all proposed similar solutions. Thus, there is still much room for research on this problem. Recently, graph neural networks (GNN) were applied to the static IM problem (Qiu et al. 2018; Tian et al. 2020; Kumar et al. 2022) and may also find success in the temporal setting.

Reverse reachable sketches

RR sketches have become the standard for state-of-the-art algorithms in the static IM setting (Tang et al. 2015; Nguyen et al. 2016; Guo et al. 2020). However, it is not straightforward to extend the RR paradigm to the temporal setting, e.g., uniformly sampling RR sets is difficult as the network evolves. RR sketch-based temporal IM algorithms have only been explored by Peng (2021), so this is an important research direction.

Ex ante vs. ex post

Most methods proposed in Sect. 2 assume that the entire topology of the dynamic network is known (*ex post* assumption), even though this is unrealistic in many situations. Yanchenko et al. (2023) yielded promising results for *ex ante* IM where the future evolution of the network is unknown, but more work is certainly needed.

Impact of time

In dynamic networks with diffusion, there is a highly intricate relationship between the structural evolution and influence diffusion. This must be carefully accounted for in the IM problem similar to Gayraud et al. (2015) and Albano et al. (2013). The impact of time scales, aggregation, diffusion times, and diffusion mechanisms deserves further study.

Online setting

Related to the previous point is IM in the online setting, where nodes and edges come and go continuously. In real-world applications, it may not be obvious how or when to aggregate the network so it becomes more natural to consider online updates. Most

methods, however, require that the network is aggregated into graph snapshots. This aggregation inherently loses information, such as when the link appeared/disappeared and the persistence of the edge. More methods like Ohsaka et al. (2016) can be developed to address this challenge.

Model mis-specification

A pertinent challenge for applied IM is selecting the diffusion model. For diseases, the SIR model is sensible since infected persons can infect other nodes for as long as they are infected. On the other hand, for HIV awareness among homeless youth, it is unlikely that someone would attempt to influence all of his/her friends indefinitely. Thus, choosing an appropriate diffusion model is crucial. But what are the effects on influence spread if the model is misspecified? In Aral and Dhillon (2018), the authors study this for static IM and find that standard diffusion models grossly underestimate the influence spread of more realistic models. This is likely only compounded in temporal networks where the topology also varies.

Uncertainty estimates of seed nodes

The majority of temporal IM algorithms output the optimal seed nodes to achieve maximal influence spread. But are there other seed sets that would yield a comparable spread? In other words, is the objective function “flat” in the sense that many seed sets yield comparable spread? An interesting avenue of research would be deriving a measure of uncertainty for optimal seed sets.

Influence minimization

A related problem to IM is that of influence *minimization* in which seed nodes are “vaccinated” to stop the spread of influence on the network. This problem arises in rumor diffusion and epidemiological settings (Wang et al. 2013, 2020; Yang et al. 2019; Wang et al. 2017) and may lead to interesting philosophical questions. For example, in the vaccine campaign against COVID-19, vaccines were first administered to the most vulnerable populations, e.g., elderly. Thus, seed nodes were chosen based on vulnerability. In an influence minimization schema, however, the most active and/or social people would likely receive the vaccine first to minimize the spread between groups. These opposing goals lead to challenging decisions both ethically and politically.

Acknowledgements

We would like to thank the editors and reviewers whose comments greatly improved the quality of the manuscript.

Author Contributions

EY was responsible for the conception, design and writing of the manuscript. TM provided editing and mentorship. PH created the original figures, and provided editing and mentorship.

Funding

This work was conducted while EY was on a fellowship funded by the Japan Society for the Promotion of Science (JSPS). PH was supported by JSPS KAKENHI Grant Number JP 21H04595

Availability of data and materials

No data was analyzed in this work.

Declarations

Conflict of interest

The authors declare no conflict of interest.

Received: 13 December 2023 Accepted: 13 May 2024

Published online: 21 May 2024

References

- Aggarwal CC (2016) Recommender systems. Springer, Berlin, vol 1
- Aggarwal CC, Lin S, Yu PS (2012) On influential node discovery in dynamic social networks. In: Proceedings of the 2012 SIAM international conference on data mining. SIAM, pp 636–647
- Albano A, Guillaume J-L, Heymann S, Grand BL (2013) A matter of time-intrinsic or extrinsic-for diffusion in evolving complex networks. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, pp 202–206
- Aral S, Dhillon PS (2018) Social influence maximization under empirical influence models. *Nat Hum Behav* 2(6):375–382
- Azaouzi M, Mnasri W, Romdhane LB (2021) New trends in influence maximization models. *Comput Sci Rev* 40:100393
- Bai Y, Yang B, Lin L, Herrera JL, Du Z, Holme P (2017) Optimizing sentinel surveillance in temporal network epidemiology. *Sci Rep* 7:4804
- Bharathi S, Kempe D, Salek M (2007) Competitive influence maximization in social networks. In: Internet and network economics: third international workshop, WINE (2007) San Diego, CA, USA, December 12–14, 2007. Proceedings 3. Springer 2007:306–311
- Bhattacharya S, Gaurav K, Ghosh S (2019) Viral marketing on social networks: an epidemiological perspective. *Phys A* 525:478–490
- Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. *Knowl Based Syst* 46:109–132
- Borgs C, Brautbar M, Chayes J, Lucier B (2014) Maximizing social influence in nearly optimal time. In: Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms. SIAM, pp 946–957
- Chandran J, Viswanatham VM (2022) Dynamic node influence tracking based influence maximization on dynamic social networks. *Microprocess Microsyst* 95:104689
- Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, pp 199–208
- Chen W, Yuan Y, Zhang L (2010) Scalable influence maximization in social networks under the linear threshold model. In: 2010 IEEE international conference on data mining. IEEE, pp 88–97
- Chen X, Song G, He X, Xie K (2015) On influential nodes tracking in dynamic social networks. In: Proceedings of the 2015 SIAM international conference on data mining. SIAM, pp 613–621
- Christakis NA, Fowler JH (2010) Social network sensors for early detection of contagious outbreaks. *PLOS ONE* 5(9):1294809
- Cohen E, Delling D, Pajor T, Werneck RF (2014) Sketch-based influence maximization and computation: scaling up with guarantees. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management, pp. 629–638
- Domingos P, Richardson M (2001) Mining the network value of customers. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining, pp 57–66
- Erkol Ş, Mazzilli D, Radicchi F (2020) Influence maximization on temporal networks. *Phys Rev E* 102(4):042307
- Erkol Ş, Mazzilli D, Radicchi F (2022) Effective submodularity of influence maximization on temporal networks. *Phys Rev E* 106(3):034301
- Feld SL (1991) Why your friends have more friends than you do. *Am J Sociol* 96(6):1464–1477
- Garton L, Haythornthwaite C, Wellman B (1997) Studying online social networks. *J Comput Med Commun* 3(1):JCMC313
- Gayraud NT, Pitoura E, Tsaparas P (2015) Diffusion maximization in evolving social networks. In: Proceedings of the 2015 ACM conference on online social networks, pp 125–135
- Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99(12):7821–7826
- Goyal A, Lu W, Lakshmanan LV (2011) Celf++ optimizing the greedy algorithm for influence maximization in social networks. In: Proceedings of the 20th international conference companion on world wide web, pp 47–48
- Grindrod P, Parsons MC, Higham DJ, Estrada E (2011) Communicability across evolving networks. *Phys Rev E* 83(4):046120
- Guimera R, Amaral LAN (2004) Modeling the world-wide airport network. *Eur Phys J B* 38:381–385
- Guo Q, Wang S, Wei Z, Chen M (2020) Influence maximization revisited: efficient reverse reachable set generation with bound tightened. In: Proceedings of the 2020 ACM SIGMOD international conference on management of data, ser. SIGMOD '20. New York, NY, USA: Association for Computing Machinery, 2167–2181. <https://doi.org/10.1145/3318464.3389740>
- Hafiene N, Karoui W, Romdhane LB (2020) Influential nodes detection in dynamic social networks: a survey. *Expert Syst Appl* 159:113642
- Han M, Yan M, Cai Z, Li Y, Cai X, Yu J (2017) Influence maximization by probing partial communities in dynamic online social networks. *Trans Emerg Telecommun Technol* 28(4):e3054
- Hao F, Zhu C, Chen M, Yang L T, Pei Z (2011) Influence strength aware diffusion models for dynamic influence maximization in social networks. In: 2011 international conference on internet of things and 4th international conference on cyber, physical and social computing. IEEE, pp 317–322
- Harush U, Barzel B (2017) Dynamic patterns of information flow in complex networks. *Nat Commun* 8(1):2181
- Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. *ACM Trans Info Syst* 22(1):5–53
- Hinz O, Skiera B, Barrot C, Becker JU (2011) Seeding strategies for viral marketing: an empirical comparison. *J Mark* 75(6):55–71

- Holme P (2004) Efficient local strategies for vaccination and network attack. *Europhys Lett* 68(6):908
- Holme P (2015) Modern temporal network theory: a colloquium. *Eur Phys J B* 88:1–30
- Holme P (2017) Three faces of node importance in network epidemiology: exact results for small graphs. *Phys Rev E* 96:062305
- Holme P, Saramäki J (2012) Temporal networks. *Phys Rep* 519(3):97–125
- Huang S, Lin W, Bao Z, Sun J (2022) Influence maximization in real-world closed social networks. arXiv preprint [arXiv:2209.10286](https://arxiv.org/abs/2209.10286)
- Karsai M, Kivela M, Pan RK, Kaski K, Kertész J, Barabási A-L, Saramäki J (2011) Small but slow world: how network topology and burstiness slow down spreading. *Phys Rev E* 83(2):025102
- Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining, pp 137–146
- Kim M, Leskovec J (2011) The network completion problem: inferring missing nodes and edges in networks. In: Proceedings of the 2011 SIAM international conference on data mining. SIAM, pp 47–58
- Kim D, Hyeon D, Oh J, Han W-S, Yu H (2017) Influence maximization based on reachability sketches in dynamic graphs. *Inf Sci* 394:217–231
- Kumar S, Mallik A, Khetarpal A, Panda B (2022) Influence maximization in social networks using graph embedding and graph neural network. *Inform Sci* 607:1617–1636
- Latora V, Marchiori M (2005) Vulnerability and protection of infrastructure networks. *Phys Rev E* 71(1):015103
- Lee S, Rocha LEC, Liljeros F, Holme P (2012) Exploiting temporal network structures of human interaction to effectively immunize populations. *PLoS ONE* 7:e36439
- Leskovec J, Adamic LA, Huberman BA (2007) The dynamics of viral marketing. *ACM Trans Web*, 1(1), 5–es
- Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N (2007) Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining, pp 420–429
- Liu W, Song Z (2020) Review of studies on the resilience of urban critical infrastructure networks. *Reliab Eng Syst Saf* 193:106617
- Li X, Du N, Li H, Li K, Gao J, Zhang A (2014) A deep learning approach to link prediction in dynamic networks. In: Proceedings of the 2014 SIAM international conference on data mining. SIAM, pp 289–297
- Li A, Cornelius SP, Liu Y-Y, Wang L, Barabási A-L (2017) The fundamental advantages of temporal networks. *Science* 358(6366):1042–1046
- Li Y, Fan J, Wang Y, Tan K-L (2018) Influence maximization on social graphs: a survey. *IEEE Trans Knowl Data Eng* 30(10):1852–1872
- López-Pintado D (2008) Diffusion in complex social networks. *Games Econ Behav* 62(2):573–590
- Lü L, Chen D, Ren X-L, Zhang Q-M, Zhang Y-C, Zhou T (2016) Vital nodes identification in complex networks. *Phys Rep* 650:1–63
- Michalski R, Palus S, Kazienko P (2011) Matching organizational structure and social network extracted from email communication. In: Business information systems: 14th international conference, BIS (2011) Poznań, Poland, June 15–17, 2011. Proceedings 14. Springer 2011:197–206
- Michalski R, Kajdanowicz T, Bródka P, Kazienko P (2014) Seed selection for spread of influence in social networks: temporal vs. static approach. *N Gener Comput* 32(3):213–235
- Michalski R, Jankowski J, Pazura P (2020) Entropy-based measure for influence maximization in temporal networks. In: Computational science-ICCS, 20th international conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part IV 20. Springer 2020:277–290
- Michalski R, Jankowski J, Bródka P (2020) Effective influence spreading in temporal networks with sequential seeding. *IEEE Access* 8:151208–151218
- Min H, Cao J, Yuan T, Liu B (2020) Topic based time-sensitive influence maximization in online social networks. *World Wide Web* 23:1831–1859
- Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In Proceedings of the 7th ACM SIGCOMM conference on internet measurement, pp 29–42
- Moreno JL, Jennings HH (1938) Statistics of social configurations. *Sociometry* 1(3/4):342–274
- Morone F, Makse HA (2015) Influence maximization in complex networks through optimal percolation. *Nature* 524(7563):65–68
- Murata T, Koga H (2018) Extended methods for influence maximization in dynamic networks. *Comput Soc Netw* 5(1):1–21
- Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions–i. *Math Program* 14:265–294
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256
- Newman MEJ (2018) *Networks*. Oxford University Press, Oxford
- Nguyen HT, Thai MT, Dinh TN (2016) Stop-and-stare: optimal sampling algorithms for viral marketing in billion-scale networks. In: Proceedings of the 2016 international conference on management of data, ser. SIGMOD '16. New York, NY, USA: Association for Computing Machinery, pp 695–710. <https://doi.org/10.1145/2882903.2915207>
- Ohsaka N, Akiba T, Yoshida Y, Kawarabayashi K-i (2014) Fast and accurate influence maximization on large networks with pruned monte-carlo simulations. In: Proceedings of the AAAI conference on artificial intelligence, vol 28
- Ohsaka N, Akiba T, Yoshida Y, Kawarabayashi K-i (2016) Dynamic influence analysis in evolving networks. *Proc VLDB Endow* 9(12):1077–1088
- Osawa S, Murata T (2015) Selecting seed nodes for influence maximization in dynamic networks. In: Complex networks VI. Springer, pp 91–98

- Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2015) Epidemic processes in complex networks. *Rev Mod Phys* 87(3):925
- Pathak N, Banerjee A, Srivastava J (2010) A generalized linear threshold model for multiple cascades. In: 2010 IEEE international conference on data mining. IEEE, pp 965–970
- Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, Schneider R, Bagos PG (2011) Using graph theory to analyze biological networks. *BioData Min* 4:1–27
- Peng B (2021) Dynamic influence maximization. *Adv Neural Inf Process Syst* 34:10718–10731
- Phuvipadawat S, Murata T (2010) Breaking news detection and tracking in twitter. In 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, vol 3. IEEE, pp 120–123
- Prakash BA, Tong H, Valler N, Faloutsos M, Faloutsos C (2010) Virus propagation on time-varying networks: Theory and immunization algorithms. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, pp. 99–114
- Qiu J, Tang J, Ma H, Dong Y, Wang K, Tang J (2018) Deepinf: social influence prediction with deep learning. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp 2110–2119
- Rodriguez MG, Balduzzi D, Schölkopf B (2011) Uncovering the temporal dynamics of diffusion networks. arXiv preprint [arXiv:1105.0697](https://arxiv.org/abs/1105.0697)
- Saito K, Nakano R, Kimura M (2008) Prediction of information diffusion probabilities for independent cascade model. In: Knowledge-based intelligent information and engineering systems: 12th international conference, KES (2008) Zagreb, Croatia, September 3–5, 2008, Proceedings, Part III 12. Springer 2008:67–75
- Shakarian P, Bhatnagar A, Aleali A, Shaabani E, Guo R, Shakarian P, Bhatnagar A, Aleali A, Shaabani E, Guo R (2015) The independent cascade and linear threshold models. Springer, Berlin
- Simmel G (1955) Conflict and the web of group affiliations. The Free Press, Glencoe, IL
- Singh AK, Kailasam L (2021) Link prediction-based influence maximization in online social networks. *Neurocomputing* 453:151–163
- Song G, Li Y, Chen X, He X, Tang J (2016) Influential node tracking on dynamic social network: an interchange greedy approach. *IEEE Trans Knowl Data Eng* 29(2):359–372
- Strogatz SH (2001) Exploring complex networks. *Nature* 410(6825):268–276
- Tang Y, Xiao X, Shi Y (2014) Influence maximization: near-optimal time complexity meets practical efficiency. In: Proceedings of the 2014 ACM SIGMOD international conference on management of data, pp 75–86
- Tang Y, Shi Y, Xiao X (2015) Influence maximization in near-linear time: a martingale approach. In: Proceedings of the 2015 ACM SIGMOD international conference on management of data, ser. SIGMOD '15. New York, NY, USA: Association for Computing Machinery, pp 1539–1554. <https://doi.org/10.1145/2723372.2723734>
- Tian S, Mo S, Wang L, Peng Z (2020) Deep reinforcement learning-based approach to tackle topic-aware influence maximization. *Data Sci Eng* 5:1–11
- Tong G, Wu W, Tang S, Du D-Z (2016) Adaptive influence maximization in dynamic social networks. *IEEE/ACM Trans Netw* 25(1):112–125
- Wang X, Deng K, Li J, Yu JX, Jensen CS, Yang X (2020) Efficient targeted influence minimization in big social networks. *World Wide Web* 23(4):2323–2340
- Wang S, Zhao X, Chen Y, Li Z, Zhang K, Xia J (2013) Negative influence minimizing by blocking nodes in social networks. In: Proceedings of the 17th AAAI conference on late-breaking developments in the field of artificial intelligence, pp 134–136
- Wang Y, Zhu J, Ming Q (2017) Incremental influence maximization for dynamic social networks. In: Data science: third international conference of pioneering computer scientists, engineers and educators, ICPCSEE (2017) Changsha, China, September 22–24, 2017, Proceedings, Part II. Springer 2017:13–27
- Wang Y, Fan Q, Li Y, Tan K-L (2017) Real-time influence maximization on dynamic social streams. arXiv preprint [arXiv:1702.01586](https://arxiv.org/abs/1702.01586)
- Wang B, Chen G, Fu L, Song L, Wang X (2017) Drimux: dynamic rumor influence minimization with user experience in social networks. *IEEE Trans Knowl Data Eng* 29(10):2168–2181
- Waniek M, Holme P, Cebrian M, Rahwan T (2022) Social diffusion sources can escape detection. *iScience* 25(9):104956
- Wilder B, Yadav A, Immorlica N, Rice E, Tambe M (2017) Uncharted but not uninfluenced: influence maximization with an uncertain network. In: Proceedings of the 16th conference on autonomous agents and multiagent systems, pp 1305–1313
- Wilder B, Onasch-Vera L, Hudson J, Luna J, Wilson N, Petering R, Woo D, Tambe M, Rice E (2018) End-to-end influence maximization in the field. *AAMAS* 18:1414–1422
- Wu X, Fu L, Meng J, Wang X (2019) Maximizing influence diffusion over evolving social networks. In: Proceedings of the fourth international workshop on social sensing, pp 6–11
- Xu B, Liu L (2010) Information diffusion through online social networks. In: 2010 IEEE international conference on emergency management and management sciences. IEEE, pp 53–56
- Yadav A, Chan H, Jiang AX, Xu H, Rice E, Tambe M (2016) Using social networks to aid homeless shelters: dynamic influence maximization under uncertainty. *AAMAS* 16:740–748
- Yadav A, Wilder B, Rice E, Petering R, Craddock J, Yoshioka-Maxwell A, Hemler M, Onasch-Vera L, Tambe M, Woo D (2017) Influence maximization in the field: the arduous journey from emerging to deployed application. In: Proceedings of the 16th conference on autonomous agents and multiagent systems, pp 150–158
- Yadav A, Wilder B, Rice E, Petering R, Craddock J, Yoshioka-Maxwell A, Hemler M, Onasch-Vera L, Tambe M, Woo D (2018) Bridging the gap between theory and practice in influence maximization: raising awareness about hiv among homeless youth. In: *IJCAI*, pp 5399–5403
- Yanchenko E, Murata T, Holme P (2023) Link prediction for ex ante influence maximization on temporal networks. *Appl Net Sci* 8(1):70
- Yang Y, Wang Z, Pei J, Chen E (2017) Tracking influential individuals in dynamic networks. *IEEE Trans Knowl Data Eng* 29(11):2615–2628
- Yang Y, Pei J (2019) Influence analysis in evolving networks: a survey. *IEEE Trans Knowl Data Eng* 33(3):1045–1063

- Yang L, Li Z, Giua A (2019) Influence minimization in linear threshold networks. *Automatica* 100:10–16
- Zhang Q, Lu J, Jin Y (2021) Artificial intelligence in recommender systems. *Complex Intell Syst* 7:439–457
- Zhou Y, Cheng H, Yu JX (2009) Graph clustering based on structural/attribute similarities. *Proc VLDB Endow* 2(1):718–729
- Zhuang H, Sun Y, Tang J, Zhang J, Sun X (2013) Influence maximization in dynamic social networks. In: 2013 IEEE 13th international conference on data mining. IEEE, pp 1313–1318

Publisher's Note

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