The b2biers System: A Content-Based Perspective on Maximizing Influence and Subscription in Social Networks

Konstantinos Theocharidis
Singapore Management University
Singapore

Hady W. Lauw
Singapore Management University
Singapore

ABSTRACT

The popular problem of *Influence Maximization* (IM) asks for the k *users* who can maximize the influence of a *fixed* post in a social network. In contrast, the problem of *Content-Aware Influence Maximization* (CAIM) asks for the k *features* to form a viral *tunable* post in a social network starting its diffusion from a *fixed* set of initial adopters. CAIM paves the way for a number of *novel problems* to be studied that altogether can lead to the development of a *system* that would be valuable for advertisers who manage social network pages. This holds since features (brands) in CAIM map to specific social network pages and each advertiser of a certain page can utilize their own feature along with others *in a variety of ways* to form a proper *content* for *influence* and *subscription* maximization purposes. In this article, we present our *content-based perspective* about how such a system (named `b2biers`) can be built, the *technical challenges* about it, and the *novel services* that it can yield to every kind of brands and advertisers running the brand pages.

1. INTRODUCTION

In this article, we present the `b2biers` system to innovatively address the *open problem* of engaging advertising in social networks. The name `b2biers` derives from the fact that different brands (features) participate in posts to achieve *influence* and *subscription*; namely, each post includes a brand-to-brand (business-to-business) collaboration.

Scope. We emphasize that this work belongs to the journal areas: (i) *Topical articles on problems and challenges* and (ii) *Well-articulated position papers* mentioned in the submission guidelines of journal. This means that we present the *design* relative to the deployment of `b2biers` system based on a plethora of realistic and diverse prior research results we achieved in [65] but we *do not* implement `b2biers` in this work. Our aim in this article is to demonstrate how our *prior research* [65] can be combined with a series of *novel services* we propose here so as `b2biers` to gradually be built and totally completed in the next few years. We have an academic background but we intend to provide our system (after completing its whole implementation) both to academy and industry communities. For that, `b2biers` will be *open-source* having *payable services* for using it.

Situation. Nowadays, most *stakeholders* (hereafter, brands) maintain pages in social networks for advertising purposes as many users having a social account opt to stay tuned with the latest news and products of a brand by following the brand page. The set of users that *follow* the page of a brand constitute the *subscribers* of brand. The person (or group) that is responsible for the *content* uploaded in the page of a brand is called the *advertiser* of brand. The goal of advertiser is to publish interesting content that is able to engage users either by *influence* (by acquiring the *like* of a user to the current post) or by *subscription* (by mainly motivating a *non-subscriber* user to subscribe to the brand page). In particular, advertiser has *viral marketing* (content that maximizes influence) and *loyalty marketing* (content that engages or increases the subscriber set) targets. The users of our `b2biers` system are such kind of advertisers.

Motivation. Till now, the *sole way* for a brand to increase the popularity of its page is to *pay the available advertising services* of social network companies. However, in practice, such services have limited options, do not apply well, they are not economic, and since their implementation details are hidden, they can also be not considered trustworthy enough. Especially for *new brands* (having a *limited* audience), such payable services request a daily and long-term budget till an adequate audience to be formed; this is an expensive and time-consuming process that is usually not preferred by those brands due to its cumbersome applicability. To address this problem, `b2biers` provides *a variety of open-source and affordable services* to the advertiser of each brand (either *new* or *established* one) to increase the popularity of their brand page. Specifically, we stress that `b2biers` is the *first system* that can *guide* advertisers in an *algorithmic* and *consistent* way to *form* engaging *content* for *influence* and *subscription* maximization purposes in social networks.

Structure. The `b2biers` system comprises a *Post Decision Engine* (PDE) mechanism (that corresponds to our recently published article [65]) and a set of associated *operations* around it that use a number of *units* to their execution. Figure 1 depicts the structure of `b2biers`. Each *operation* combined with PDE generates a new *service* provided by `b2biers`. The PDE of `b2biers` applies a *reinforcement learning* method to adaptively *eliminate* features found less influential than others over a series of rounds. Yet, in each round, k *non-eliminated* features are *randomly selected* to form the *post* of current round. Instead of *random selection* that takes place in [65], a variety of *operations* can be utilized in `b2biers` for the creation of round *post*, each one of them contributing to a different *influence* or *subscription* maximization target. Still, as the execution of `b2biers` is continuous, when a large portion of features are eliminated, there would be a *revive* mechanism that brings back to life

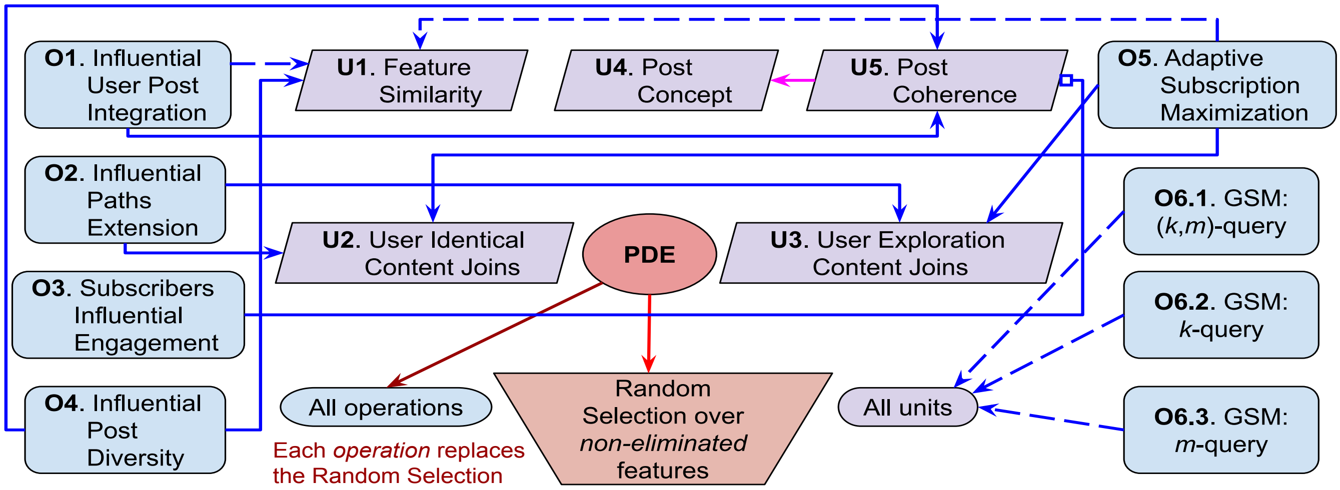


Figure 1: The *operations* (blue) and *units* (purple) of the `b2biers` system. Each operation uses a number of units but does not use another operation. The *advertiser* of relative *brand* selects in each round **strictly** one *service* (meaning one operation from the PDE of `b2biers`) to replace the *random selection* (taking place in [65]) so as to utilize the *post* of round for *influence* or *subscription* maximization goals. The symbol → means use, ⇢ means optional use, and □ means use all except for this.

eliminated features in a gradual way based on their importance (we omit *revive* mechanism from Figure 1 for clarity; this mechanism is not implemented in [65]).

PDE. The PDE of `b2biers` constitutes its *central mechanism* because it relates with a set of *non-eliminated* features via which k features should be selected to form the post of round. Each *feature* corresponds to a specific *social network page* that has a number of *subscribers* who follow the page. A feature can be any kind of page; e.g., *Laughter is the best medicine*, *Charlize Theron*, *Rome* are all features. The best *learner* we developed in [65] eliminates features that found, based on a simulation-feedback, that are less influential than others. So, PDE selects even more influential features over rounds and this process helps to maximize the cumulative influence over rounds (that is the target in [65]). Although that feature-selection is *randomly* done in [65], it performs well for a variety of brands, and that motivated us to leverage that PDE learning-mechanism for our `b2biers` purposes. The light-red-arrow in Figure 1 maps to the execution of PDE as happens in [65], and the dark-red-arrow maps to the execution of PDE as happens in `b2biers`.

Services. Each *service* of `b2biers` achieves an *influence* or *subscription* maximization goal. Maximizing influence relates with the popular problem of *Influence Maximization* (IM) [35; 47] that searches for the k users who maximize the influence (number of *likes*) of a *fixed* post in a social network. Yet, every influence target of `b2biers` relies on an *inverse* direction of IM, first introduced in [33], which relates with maximizing influence by finding a set of k *features* (content) that form a viral *tunable* post. In the *Content-Aware Influence Maximization* (CAIM) problem of [33], each *feature* maps to a specific *social network page*, and each post propagation starts from the *subscribers* of brand page. Searching and combining such kind of *features* to achieve influence can really derive a broad and creative research with different goals that is able to lead to an interesting system like the `b2biers` system we present in this article. Further, such kind of *content-based influence* more naturally leads to *subscription* maximization or gaining services to be deployed since gaining the subscription of a user can be modeled as the *repetitive* content-aware influence on user.

The main contributions of this article are the following:

- **System.** We present our *design* relative to the development of the `b2biers` system. In contrast to the *commercial* and *limited* services provided by the social network companies to brands to advertise their pages, `b2biers` supports *a variety of open-source* and *affordable* services for *influence* and *subscription* maximization goals. `b2biers` depends on our prior research [65] (mentioned as PDE in this article) by proposing a series of *innovative operations* that utilize PDE.

- **Technical Challenges.** We introduce 5 *units* and 8 *operations* depicted in Figure 1 to form `b2biers`. Each one of such 13 *components* we propose are *independent* and *novel* problems having their own distinct technical challenges, besides the fact that we scheduled them to utilize the PDE component we implemented in [65].

- **Applicability.** We discuss several scenarios and examples relative to the applicability of `b2biers` in real world. Each one of them relates with the *novel components* (units and operations) we propose in this article.

- **Related Work.** We present a *broad* and *detailed* related work to emphasize and clarify that `b2biers` is *innovative* under both *academic* and *industry* terms. Particularly, we first discuss how each one of proposed 5 units and 8 operations (in short, referred as *components*) relates with literature to justify that each component really depicts a *novel problem.* We also illustrate how advertising takes place in the social network industry by discussing advertising mechanisms that utilized by the most popular social networks.

2. PDE OF b2biers

In this section, we present the *Post Decision Engine* (PDE) mechanism of the `b2biers` system. By the term PDE we

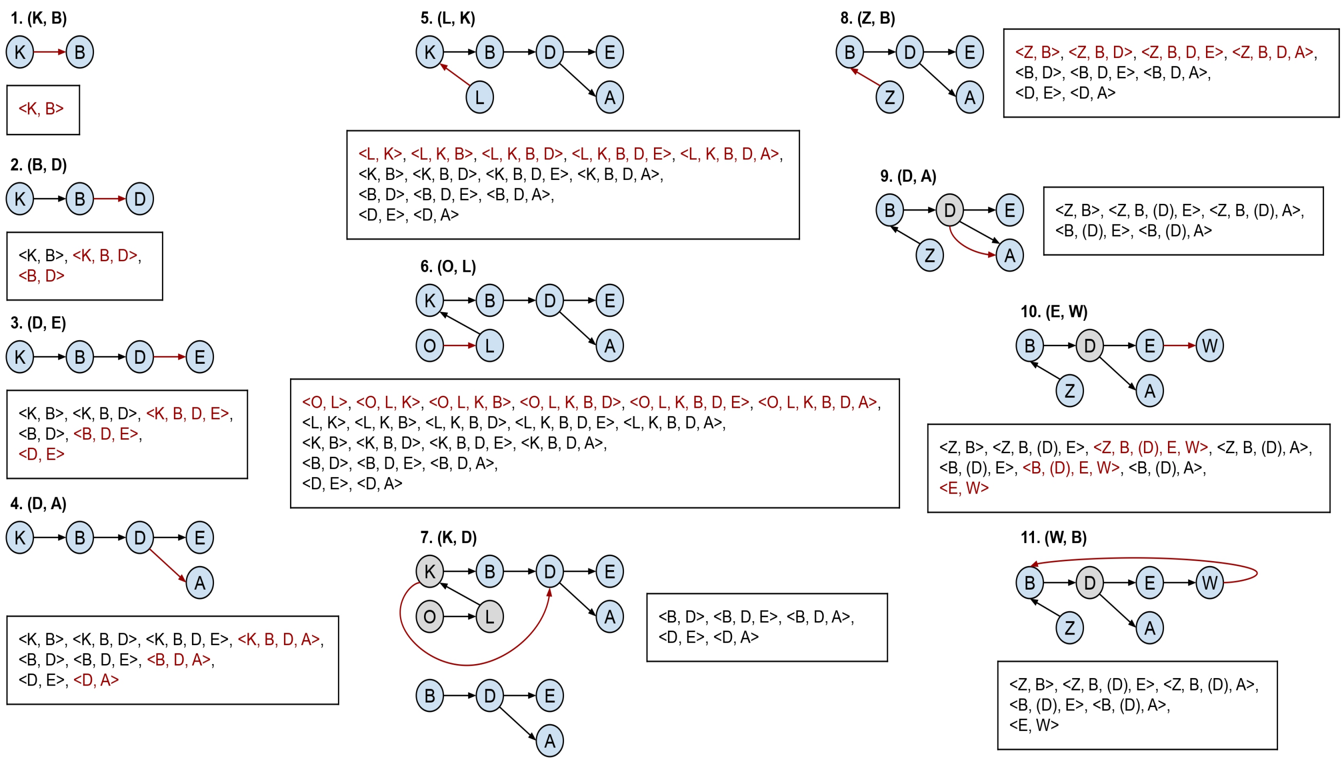


Figure 2: Flow of TRIM_E. Each step digests new evidence (feature path, *red*); in step 7, (K, D) eliminates features in grey; in step 9, the eliminated D remains as intermediate; in step 11, the cycle caused by (W, B) cancels paths containing subpath $\prec$B, (D), E, W$\succ$. At each step (f_1, f_2), a feature f_1 that is found less influential than another feature f_2, incurs changes to $\mathcal{T}$; these changes depicted as the *red* set of paths in respective tables. This Figure is the **exact copy** of Figure 3 in [65].

Table 1: Percentages (%) depicting how much ***influence better*** TRIM_C and TRIM_E are over RANDOM for $\mathbb{A}$ and $\mathbb{B}$. This Table is the **exact copy** of Table 4 in [65].

Algorithm	RANDOM (A1, B1)			RANDOM (A2, B2)			RANDOM (A3, B3)			**Average**		
	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$
TRIM_C ($\mathbb{A}$)	23.3	71.2	123.9	15.6	48.1	78.9	43.7	130.5	171.1	27.5	83.2	124.6
TRIM_C ($\mathbb{B}$)	8.6	320.9	849.7	3.3	15.3	32	2.1	6.2	108.6	4.6	114.1	330.1
TRIM_E ($\mathbb{A}$)	46.5	97.9	154.3	23.6	58.1	95.4	67.9	177.3	224.4	46	111.1	158
TRIM_E ($\mathbb{B}$)	329	1372.9	1646.9	5.6	26.5	45.8	32.9	148.9	229.4	122.5	516.1	640.7

Table 2: Percentages (%) depicting how much ***influence better*** TRIM_E is over TRIM_C for $\mathbb{A}$ and $\mathbb{B}$. This Table is the **exact copy** of Table 5 in [65].

Algorithm	TRIM_C (A1, B1)			TRIM_C (A2, B2)			TRIM_C (A3, B3)			**Average**		
	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$
TRIM_E ($\mathbb{A}$)	18.8	15.5	13.5	6.9	6.7	9.2	16.8	20.3	19.6	14.1	14.1	14.1
TRIM_E ($\mathbb{B}$)	294.7	249.9	83.9	2.1	9.6	10.4	30.1	134.3	57.8	108.9	131.2	50.7

Table 3: Percentages (%) depicting how much ***learning faster*** TRIM_E is over TRIM_C for $\mathbb{A}$ and $\mathbb{B}$. This Table is the **exact copy** of Table 6 in [65].

Algorithm	TRIM_C (A1, B1)			TRIM_C (A2, B2)			TRIM_C (A3, B3)			**Average**		
	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$
TRIM_E ($\mathbb{A}$)	-2.2	219.1	104.7	-25.8	22.2	328	1.1	58.4	551.3	-8.9	99.9	328
TRIM_E ($\mathbb{B}$)	> 17.1	259.2	448.5	> 6.2	89.2	146.9	> 20.7	> 295.2	> 601.7	> 14.6	> 214.5	> 399

mention to our prior work found here [65]. In that work, we implemented three learners named RANDOM, TRIM_C, and TRIM_E, but as shown there, TRIM_E outperforms significantly the others, so we focus our discussion here only on the execution of TRIM_E as this is the learner we keep for our `b2biers` system; the learners RANDOM and TRIM_C are not part of `b2biers`. In the following, we discuss the basic execution of PDE for completeness; for further details, readers could consult our prior work [65].

The PDE consists of a *learner* and a *simulator*. In each round, learner forms a post comprising k features (the *brand* feature along with $k-1$ features depicting *other* brands) and receives a *feedback* from simulator; feedback is defined as the set of users who *liked* the propagated post of round. The objective of learner is to maximize the *cumulative* influence spread (number of total *likes*) over all rounds. To achieve that, for each mentioned feedback user, learner estimates which are the post features that motivated the user to like the post. Guided by that estimation, learner assigns a *conceptual click* to each feature that stimulated the interest of every feedback user. Namely, learner finds the total number of clicks aggregated by each feature of the post after considering all feedback users.

The more clicks collected by a feature the more influence it generates to the network. So, learner utilizes a *transitive* structure $\mathcal{T}$ that stores and manages *influence comparison paths* among features in a *transitive* way to find influential features over rounds. For instance, suppose that after processing the feedback of *first* round, learner inserts to $\mathcal{T}$ the path $\prec a, b \succ$. This means that there is a *suspicion* that feature a is *less influential* than feature b. That suspicion derived since a did not collect enough clicks compared to b by considering also the *different portions* that a and b have in the post of *first* round; the k features of any post have different portion based on their importance (*brand* feature always has the biggest portion). Assume that the path $\prec b, c \succ$ inserted to $\mathcal{T}$ in *second* round, then $\mathcal{T}$ contains three paths $\{\prec a, b \succ, \prec b, c \succ, \prec a, b, c \succ\}$. The path $\prec a, b, c \succ$ derived in a *transitive* way and means that there is a suspicion that a is less influential than c. In case that a *repetitive* suspicion is found among connected features, then the source feature is *eliminated*. E.g., in the presence of a path $\prec a, c \succ$ in *third* round, the content of $\mathcal{T}$ is $\{\prec b, c \succ\}$ as feature a eliminated and that also removed from $\mathcal{T}$ any path in which a appears as source or destination for memory saving reasons. Instead, in the presence of a path $\prec c, b \succ$ in *third* round, the content of $\mathcal{T}$ is $\{\prec a, b \succ\}$ since an *opposite* suspicion (*cycle* event) among b and c is found and so any knowledge depending on $\prec b, c \succ$ is canceled. Figure 2 presents an execution flow of learner TRIM_E using $\mathcal{T}$ in which *elimination* and *cycle* events can be detected.

Both learner and simulator initialize their parameters based on the same *training phase* (a big set of realistic posts considering each post published in page of each feature). The difference is that in *testing phase*, the learner has *partial* knowledge of the propagation model (that knowledge relates with the parameters tuned in training phase), while simulator has *complete* knowledge of the propagation model (additional parameters besides the ones tuned in training phase). In particular, the learner does not know how much a user should be affected to eventually like (influenced by) the propagated post and also the learner is not aware of any of the posts in testing phase that affect the influence decisions of users. Such things are known *only* to simulator, which makes use of *realistic posts* in testing phase to decide the feedback (influenced users) of propagated post. The *trustworthy* execution of simulator leads to a *meaningful learning* for business purposes (as advertisers can really populate their pages by using the PDE of `b2biers`).

In each round, learner *randomly* selects $k-1$ *non-eliminated* features to participate in different ranks (portions) of current round post. Based on the feedback of each round it updates $\mathcal{T}$ and when a large number of features is eliminated, it brings back to life features that can be influential based on current statistics of network relative to the popularity of a revived feature (the *revive* mechanism will be provided by `b2biers`; it is not implemented in [65]). Independently of the target (maximize *influence* or gaining *subscription*), the goal of learner is to find *influential features* over rounds. Yet, the *random selection* of learner over *non-eliminated* features (that happens in [65]), creates the chances for *a variety of operations* (as earlier mentioned) to be deployed that can achieve influence or/and subscription under different scenarios. Namely, we intend in `b2biers` each *operation* to replace the *random selection*, and along with PDE to create a valuable *service* for current maximization goals. We presented in detail all such *services* of `b2biers` in Section 3.

The experimental evaluation of PDE was done on a plethora of different and realistic *case studies* (brands) in the social network VK[1]. VK has in total 27 *categories* and we selected the 10 and 20 most popular features from each category to form data $\mathbb{A}$ (contains 270 features) and data $\mathbb{B}$ (contains 540 features). Then, we created scalable datasets A1, A2, and A3 belonging to $\mathbb{A}$ and scalable datasets B1, B2, and B3 belonging to $\mathbb{B}$. All datasets have millions of nodes (users) and edges (user connections). As dataset number or letter increases, the nodes and edges increase (e.g., A2 is larger than A1 and B2 is larger than A2). Also, each dataset from the mentioned six ones, corresponds to a different brand (*case study*) that solves the PDE problem. There are a lot of experimental results that readers can find in our PDE work [65]. Here, we simply present some indicative results to justify the superiority of TRIM_E over other learners.

Table 1 shows the superiority of the learners TRIM_C and TRIM_E over the learner RANDOM in terms of *influence spread*; it mentions to the cumulative number of *likes* the posts receive over 2000 rounds (all results in [65] pertain to 2000 rounds). Table 2 shows that TRIM_E achieves a clearly higher influence than TRIM_C, especially in data $\mathbb{B}$ the influence gap is big, a fact that justifies that TRIM_C is not scalable to network size. Finally, Table 3 shows the superiority performance of TRIM_E over TRIM_C in terms of *learning speed*; it mentions to the round where *elimination* of features is over (the $k-1$ most influential features for the brand are found). Overall, these results show that a fast and accurate learning leads to higher cumulative influence spread over rounds; the most representative fact is for $k=4$ in data $\mathbb{B}$, where TRIM_E is 131.2% influence better and at

[1]VK (`https://vk.com/`) represents the Russian version of Facebook in terms of scale, functionalities, variety of topics, user accounts, brand pages, etc. It has a much more flexible and unrestricted API (`https://dev.vk.com/en/reference`) than rest social networks. Further, according to *Wikipedia*, VK had been the 16th most visited website in the world and at the moment it has more than 800M users. So, VK is very suitable as a social network data source for research aims.

least 214.5% learning faster than TRIM_C.
Finally, we stress that in [65], the learner TRIM_E was totally evaluated on 45 *case studies* of VK social network; as mentioned, each case study corresponds to a different brand (feature). The impressive performance of TRIM_E in all such realistic and diverse *case studies* convinced and motivated us to feasibly design the development of `b2biers` system that we present in this article. Readers are highly encouraged to see the results in [65], a subset of which are reproduced here for self-containment reasons.

3. SERVICES OF b2biers

In this section, we present the *services* designed to be provided by `b2biers`. As mentioned, each *service* is a combination of PDE (discussed in Section 2) and *operation(s)* (discussed in this section). All the *operations* use a number of *basic units* to their execution. So, before describing *operations*, we first illustrate and discuss the *units* taking place in their execution, which are crucially important having their own remarkable technical challenges. For simplicity, consider k as $k-1$ in the rest of article apart from certain cases or other places where we specifically discuss $k-1$ features.

3.1 Units

U1. Feature Similarity. The unit of *feature similarity* takes a pair of features and finds how similar the features are. We remind that each feature maps to a specific social network page, so feature similarity is identical to page similarity; each page is treated as a brand since it has a specific number of subscribers. We define feature similarity as a *multidimensional similarity join* on the preferences of feature subscribers. A subscriber from one feature can be matched (joined) with only a *single* subscriber from the other feature. Similarity equals to the *ratio* of number of matches to the capacity of smaller number of subscribers among two features. A user is matched with another user only if the *absolute delta* of their preference weight per dimension (each dimension maps to a weight-value of a different category) is below a threshold ϵ, and this should hold for every dimension; ϵ is *as minimum as possible* to really find similar users.
The *technical challenge* of the unit U1 is the *fast* execution of multidimensional similarity join for *every* pair of *non-eliminated* features. Similarity of features is very important for a variety of tasks, and since user preferences change often over time, efficient execution is crucial. Further, the feature similarity can be computed in an *approximate* or *exact* way, and the latter is clearly more time-consuming as we have shown in [66]. Specifically:
The *first version* of similarity join operator U1 for a *single* pair of features with *static* user preferences is implemented in our prior work [66]. Yet, the realistic nature of `b2biers` requests a similarity join operator over *all pairs of features* with *dynamic* user preferences, and implementing an efficient version of that is not a trivial task.
U2. User Identical Content Joins. The unit of *user identical content joins* finds all pairs of users (given a set of users) that like identical posts (content) over a specific time period. In more detail, for every user v it is sufficient to join v with m other users who like at least k identical posts over a time period tp depicting a number of days.
Such a unit is useful since it can connect *similar-mind* professionals/people that advertisers can exploit for the aims of `b2biers`. Joins of that kind could be models and photographers, actors and directors, household-cooks and chefs, travel fans and travel bloggers, and so on. To be more specific, consider the following real-world example:

EXAMPLE 1. Yorgos Lanthimos *is a director and* Emma Stone *is an actress. Both they are popular and successful in the worldwide cinema. They have developed a very close relationship expressed by their cooperation in several movies (depicted in Figure 3; there is also a movie of them named "Kinds of Kindness" played on cinemas in* 2024*). Our join* U2 *could connect users such as* Yorgos *and* Emma *since they would like* ***identical*** *posts in the relative social network.*

The *technical challenge* of U2 is to *quickly* apply the join for *every* user v. We argue that this can be done by a *reinforcement learning* way starting from current time and going back in time (as long as tp dictates in the worst case). The goal is to examine the *minimum* portion of posts published over tp so as to find the m matches for every user v. For that, an *exploration-exploitation* scheme is needed to guide the search space over examined posts; e.g., we could first find a set of frequent likers by exploration and then exploit that set to minimize the number of examined posts.
The unit U2 is different from the *feature similarity* unit U1. The latter, although strict, is more general as it examines all the dimension preferences of users (each dimension aggregates the *user likes* to posts published by brands belonging to specific category) to find whether two users can be matched. Namely, it checks whether two users have *similar-profile* whereas U2 checks whether two users have *similar-mind*. The idea of *similar-mind* to find similarity among two features would be *too strict* to be applied in practice, and so we consider it conceptually prohibitive for U1.
U3. User Exploration Content Joins. The unit of *user exploration content joins* finds all pairs of users (given a set of users) that have at least p_1% *similarity* and p_2% *dissimilarity* on the posts (content) of a specific category cg; $p_2 = 100 - p_1$. Specifically, similar to U2, it is sufficient to join each user v with m other users who satisfy the referred similarity-dissimilarity condition for k features (pages) of category cg. We compute similarity and dissimilarity based on the ϵ-idea of *feature similarity* unit (U1). In particular, an *absolute delta* below and above ϵ depicts similarity and dissimilarity, respectively. Also, note that the join of this unit *solely* focuses on the features (pages) belonging to category cg. Yet, as each cg can include many pages in social networks, the join process relates with several features, and so it is a kind of *multidimensional similarity join* (each dimension maps to a weight-value of a different feature).
The higher the similarity the more valuable can be the *exploration* of dissimilar pages, but if similarity is high enough then the search space of dissimilarity gets limited. A balanced value for p_1 and p_2 is 50% but their tuning is set by advertiser based on their goals. This unit is called an *exploration join* as the next example shows:

EXAMPLE 2. *Suppose a user* v_1 *has* 60% *similarity and* 40% *dissimilarity with another user* v_2 *for* cg = *"Fashion". This means that for the associated* $k = 10$ *pages of "Fashion",* v_1 *and* v_2 *similarly like* 6 *of them whereas they like in different weights the rest* 4 *of them. Still, this does not*

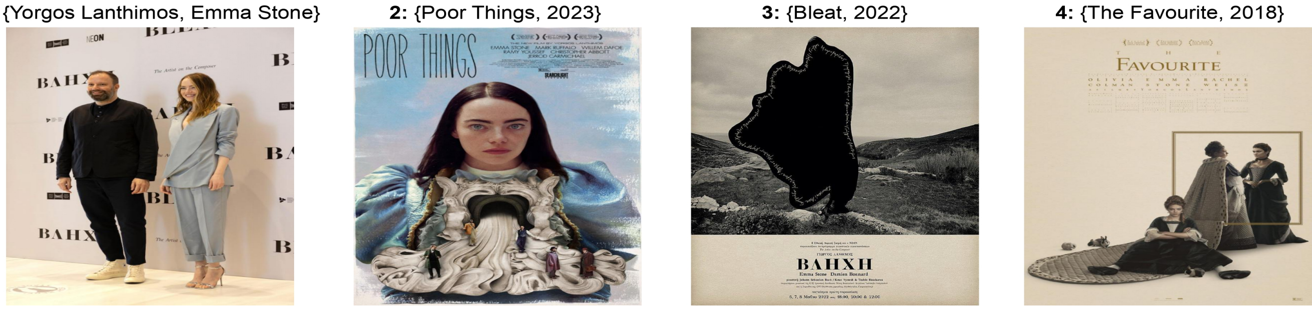


Figure 3: Yorgos Lanthimos and Emma Stone along with their *partnered* movies. Pictures taken from `https://imdb.com/`.

mean that v_2 does not really like the 4 pages that v_1 likes; this may happen because v_2 did not pay attention to them for other reasons irrelevant to the content of those pages. So, our join with $p_1 = 60\%$ and $p_2 = 40\%$ will connect v_1 with v_2 and would help an advertiser (who manages a page about "Fashion" having v_1 among subscribers) to ***explore*** *the 4 pages that v_2 likes more than v_1 and form a content based on them. By doing that, advertiser not only publishes non-repetitive (diversified) content to their page suitable to the page audience but also easily stimulates the interest of v_2 to subscribe to their page after messaging v_2. Such cases cannot be covered by prior user-identical-content-joins (*U2*) since the users v_1 and v_2 in most cases will not be connected by the execution of* U2*.*

The *technical challenge* of the unit U3 is the same as the unit U1. Namely, a really *fast* execution is needed as the user preferences (that affect similarity and dissimilarity scores) frequently change over time. Advertiser should be able to *quickly* have available the exploration join results relative to *non-eliminated* features of *cg* as time-distance among rounds can be just few hours. We also stress that the unit U3 can be approached in a *reinforcement learning* way similar to the one we mentioned for *user identical content joins* unit, and that increases further its technical challenges.

U4. Post Concept. The unit of *post concept* takes a post *pt* and finds the *k keywords* that mostly describe it (form the concept of *pt*). This unit assumes that the *only* information available for *pt* is the brand that published it and the set of users that liked *pt*; the *text description* of *pt* is ignored since usually in social networks such a description is very vague or completely absent. So, this unit exploits the *preferences* of users who like *pt* and *optionally* their social connections along with their preferences, but in most cases U4 handles only the users who liked *pt* to find its concept.

The approach to solve U4 is to first consider that only *few brands* (over all features in our social network) can be described in a *ground truth* way via sources such as *Wikipedia*. By describing, we mean the assignment of specific and realistic keywords to describe such brands; e.g., the brand *Dior* via Wikipedia is described with keywords *luxury*, *fashion*, *cosmetics*, *perfumes*, and others. Then, the goal of U4 is to guide the search process over the set of users who liked *pt* in a way that can find more of the mentioned *ground truth* so as to characterize each user and they by their turn to characterize *pt* (defining its concept). The characterization of a user is done by finding their top-k keywords (the ones with the highest occurrence) and likewise the concept of a post is defined by its k most popular keywords (derived from the user-keywords aggregation over users who liked that post). The search process of U4 is executed in a *recursive way* since to characterize a user relative to a feature, we need to examine a *specific* number of recent posts published by the feature and liked by the user in *the frequent case* that the feature will not be related with a *ground truth*, and this means that we need to solve again U4 for each one of the referred feature posts to approximate their concept via the recursive search process instead of skipping them. The top-k keywords over examined concepts via aggregation, define the k keywords of mentioned user relative to above feature. This process is repeated for each other social network feature that published a *specific* number of recent posts that user liked in order the top-k keywords that characterize the user to be found after feature-keywords aggregation. Overall, it is expected that, the higher the recursion depth we allow in our search process via a relative parameter, the higher the derived post concept accuracy of original post *pt*. Yet, higher recursion depth also leads to higher execution time, so the discussed search process is not a trivial problem. Last, we stress again that all the keywords we mentioned are part of a *ground truth* that relates only with *few brands*.

The unit U4 is important to advertisers for several reasons. E.g., they can understand much better what kind of posts their audience (subscribers or influenced users) prefer and utilize the derived information for achieving subscription and influence to other users (different than the usually affected ones). Another use is, advertisers to examine if some k features selected to form the post of current round, can really correspond to an *adequate* number of successful posts having similar concept with the one depicted by k features; if not, some other k features are selected, and so on.

We state that *preferences* and *keywords* differ. The former map to specific social network pages (considered as *features* in `b2biers`). The latter correspond to single words such as *expensive*, *nature*, *children*, etc. To illustrate how keywords can express the concept of a post, we provide the following real-world example:

Example 3. *Figure 4 shows four pictures depicting four* BMW *cars, each one relative to a different concept expressed by a set of $k = 3$ keywords; we selected the keywords and we believe they are realistic enough. We did such a selection as no picture has a sufficient description in their relative posts (the caption of figure says how to access each picture) to capture their concepts. This is the problem we intend our* ***post concept*** *unit to solve when implemented in practice.*

The *technical challenge* of the unit U4 is to find an *efficient* and *effective* method to yield the concept of *pt*. Ef-

Figure 4: Four BMW cars with *keywords* capturing the different *concept* per picture. All pictures are taken from the official page of BMW in Instagram (`https://instagram.com/bmw/`) and are accessible with prefix `https://www.instagram.com/p/` followed by `CvA7ZweK23U/` for 1, `CrmQQ3BK2xy/` for 2, `CqLnc_QIbfA/` for 3, and `COoYagvqRhf/` for 4.

Figure 5: Instagram pictures of Pom Klementieff (`https://instagram.com/pom.klementieff/`) with *keywords* expressing their *concept*; *other* refers to one other word of *concept*. Each picture is accessible with prefix `https://www.instagram.com/p/` followed by `CM56ME6lQOH/` for 1, `CpJ_Ykbryns/` for 2, `C07l5Hprz-z/` for 3, and `C2ItnXnyrs9/` for 4.

ficiency relates with the *minimum* possible examination of users (along with the processing of their preferences corresponding to social network features). Effectiveness relates with the *maximum* possible concept accuracy of pt derived from the found k keywords. Still, we stress that in practice the post concept unit should be executed for a big number of posts, and that enhances further its technical challenges. Finally, we reckon that a *reinforcement learning* approach (*exploration-exploitation* scheme) may be a good solution to also address the unit U4 as earlier mentioned ones.

U5. Post Coherence. The unit of *post coherence* checks if some given k features (pages) can actually form a coherent post. To do that, it uses the *post concept* unit (previously described) over a variety of posts. Specifically, for each one of the k features separately, it uses the *post concept* unit for a *fixed* number of feature posts (starting from current time to past time) till finding a *significant overlap* among the concept keywords (single words) relative to each one of the k features to other ones. Based on that, the post coherence unit outputs an *overlap score* for the given k features. The higher the *overlap score* the more *coherent* is the relative post of given k features. In more detail, we present the next example to precisely explain how *significant overlap* and *overlap score* are computed:

Example 4. Pom Klementieff *is an actress that often takes place in action movies like the recent "Mission Impossible". Assume that the last* 5 *posts in* BMW *and* Pom *Instagram pages are the four ones depicted in Figures 4 and 5, respectively, plus one other random post in both cases. Suppose that a keywords overlap is significant if it corresponds to at least* 50% *of common keywords. We see that among the four shown posts in both figures (the ordering could be different but one post compares with only one other post), the overlap is significant as it maps to* 2/3 *common concept keywords for each post among* BMW *and* Pom. *Also, let the overlap between random posts be* 1/3 *(not significant). So, the overlap score equals to* $4/5 = 80\%$ *since* 4 *out of* 5 *compared posts yield a significant overlap. Another popular actress like* Marion Cotillard, *for the same mentioned* BMW *posts, would probably have an overlap score close to* $20\% - 40\%$ *as she can suit only to "luxury" and "family" concepts but not to "old-school" and "adrenaline" ones. The conclusion is that* Pom *(representing competitive candidates) and* BMW *are much more coherent than* Marion *(representing normal candidates) and* BMW, *so* BMW *would select* Pom *to make a post with her; this is what can be found if* BMW *would use our* ***post coherence*** *unit when implemented in practice. To illustrate further the realism of this example, we stress that* BMW *and* Pom *have recently cooperated in a big video campaign*[2] *for* BMW *and we believe that this video*[3] *(depicting* Marion *to advertise* Dior*) indicates what* Marion *expresses in a very representative way. Note that the posts in Figure 5 also request a* ***post concept*** *unit, since none of the pictures can be concept-characterized after visiting the relative Instagram posts of* Pom.

Finding coherent posts is very important for advertisers, because the *more coherent* are the k features (higher *overlap*

[2] `https://youtube.com/watch?v=TJfAOBk7HgQ`
[3] `https://youtube.com/watch?v=UXEbtqU_dHs`

score) the *more easy* is for advertisers to form the post of current round that comprises those k features. So, we emphasize that the *coherence* of k features is not only related with the *naturalness* of the formed post but also with *how fast* that post can be created by the advertiser.

The *technical challenge* of the unit U5 is the *efficient* computation of *overlap score* that also depends on *how efficient* is the *post concept* execution of the unit U4. There is a need for a *concurrent mechanism* that updates the *overlap score* over the posts associated with k features gradually as we proceed back in time. At the same time, processing of posts should be skipped when there is found *evidence* (from prior similar posts) that such kind of posts cannot update the *overlap score*.

Another interesting variation of post coherence unit is to simply check whether the given k features are coherent or not without finding the precise *overlap score*; e.g., the features are coherent only if the *overlap score* is at least 75%. Such a problem setting may exploit better, *reinforcement learning* approaches discussed in prior units.

3.2 Operations

We remind that *influence* relates with gaining the *like* of a user to the formed post whereas *subscription* relates with gaining the subscription of a *non-subscriber* user to the brand page. Although some of the *operations* can be tuned to achieve both *influence* and *subscription*, for clarity, we separately discuss four operations for *influence* and another four operations for *subscription*. Moreover, we stress again that the role of each *operation* is to replace the *random selection* over *non-eliminated* features that takes place in [65] (discussed in Section 2) for the purposes of `b2biers`.

We remind that advertiser should select **strictly** one *operation* to execute in each round from the PDE of `b2biers` as shown in Figure 1. It is possible for advertiser to select more-than-one operations for multi-objective goals (e.g., influence some users and at the same time gain the subscription of some others), yet we emphasize that such kind of multi-objective decisions are out-of-the-score of this article.

O1. Influential User Post Integration. The operation of *influential user post integration* searches for a suitable influential user to take place (as an existing picture of them) in the post of current round based on some *criteria* that the user should fulfil. Such criteria are the user (i) to like posts having a concept of *significant overlap* with the post concepts relative to the rest $k-1$ features of the post, (ii) to have a *subscription* as also a *high preference* to the rest $k-1$ features of the post, (iii) to be *popular enough* (having a number of friends and followers above a given threshold). This operation uses the *post concept* and *post coherence* units, and it optionally uses the *feature similarity* unit to enlarge the search process by features that are similar to the k features of post (excluding the participated user).

Integrating the *proper influential user* to a post can increase the achieved influence of post. This happens due to *personalization effects* incurred by participated user to their own audience and the interesting *creativity* expressed by the post itself. The *criteria interpretation* is that chances are to find a *good picture* of user if user likes things under *similar perspectives* and of *similar aesthetics* relative to the other $k-1$ features of the post. If does that, and likes much those features separately, and also is popular, then chances are that user often uploads pictures (*depicting the user*) that can be good candidates to take place in the post of round.

The *technical challenge* of O1 is the many users to be examined along with their criteria over all the k-out-of-L-feature-combinations where L denotes the set of *non-eliminated* features. There could be an *efficiency-effectiveness* tradeoff to address this problem in case an *approximate* solution is adequate; else, *efficiency* is the sole requirement for an *exact* solution. This problem can also be addressed in *many rounds*, where in each round a different user (with a picture format) participates in current *collage post* and the aim is to maximize the *cumulative influence* over all rounds.

CASE 1. *Suppose an advertiser that maintains a brand page named* W-Fashion *relative to* woman fashion *and in current round they want to apply* personalization effects *to make more familiar the brand* W-Fashion *to their audience; so, advertiser selects from the* PDE *of* `b2biers` *the operation* O1. *Assume $k = 4$ features, hence besides* W-Fashion, *three features are searched to form the post of current round where one of them will correspond to a social network user as* O1 *dictates.* Figure 6 *depicts such a scenario in which advertiser selects the features* Travel is a pleasure *and* Starbucks *from the non-eliminated set and the user with an* entrepreneur profile *to integrate to a post that is relevant to* W-Fashion, Travel is a pleasure, *and* Starbucks. *Note that this process is just an instance of* O1*; namely, advertiser should form the post of current round* ***after*** *executing the process in* Figure 6 ***for all*** *the possible features in the non-eliminated set and their respective social network users.*

O2. Influential Paths Extension. The operation of *influential paths extension* selects the k features over the *non-eliminated* feature set that can increase the *length* of influential paths among subscribers of advertised brand and their further non-subscriber connections in network. Namely, the target is to form influential posts that can extend the *average number of hops* among subscribers and non-subscribers as rounds evolve. To do that, O2 uses the aforementioned *content joins* U2 and U3. E.g., if a subscriber set S_1 influenced in previous round, then we apply the prior *content joins* to S_1 to find *feasible influence targets* connected to S_1 but not influenced in previous round.

The value of this operation is high since its successful execution opens the way for a *fruitful engagement* of brand with a good amount of users connected (via 1-hop, 2-hop, 3-hop, etc.) with subscribers. Such *non-subscriber* users can help the brand to increase further the length of its influential paths and more importantly they can constitute its *first feasible targets* for gaining their subscription.

The *technical challenge* of the operation O2 is that there are several ways to extend the paths from subscribers to the rest of the network. Also, influence of subscribers is not taken for granted, so actually, each influence path starts from the advertised brand itself. Further, as earlier stated, there are many k-feature-combinations to be checked, as also user preferences change often over time and that affects the execution of mentioned *content joins*. Selecting the proper k features to form the post of current round that contributes to an average hops-extension is not a simple problem.

CASE 2. *Advertiser of* W-Fashion *observes that so far* mostly *subscribers of* W-Fashion *like the posts published in the brand page, and for that, advertiser wants to* expand *the*

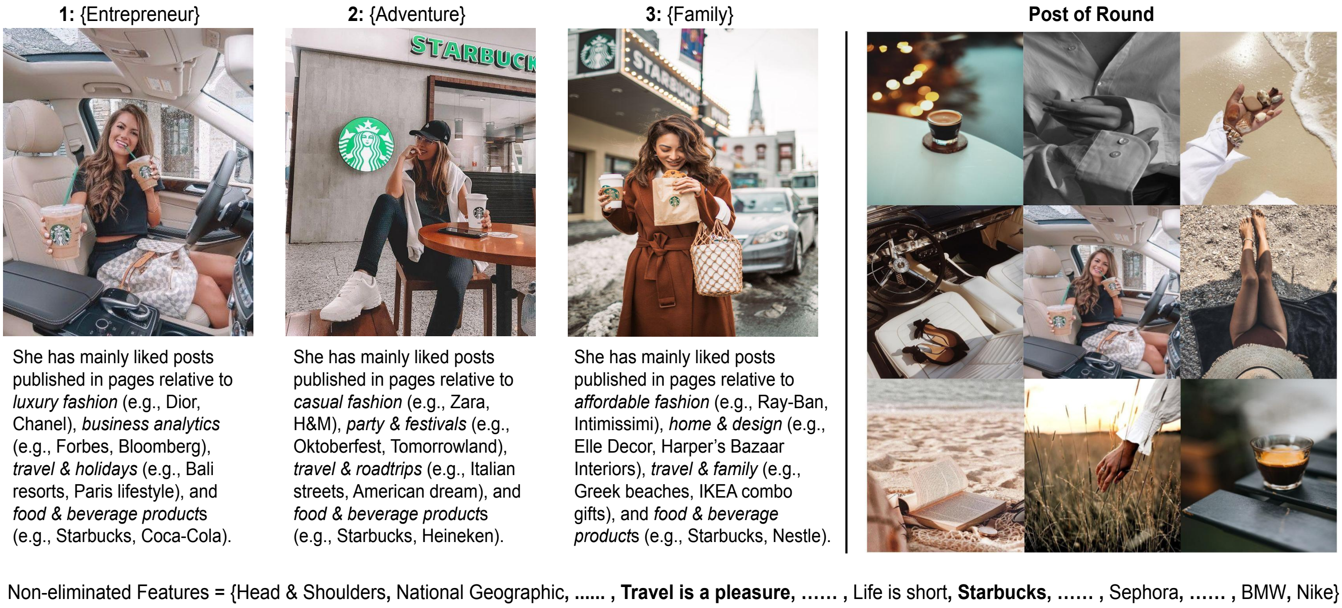


Figure 6: During the examination process of *non-eliminated features* for O1 goals by the advertiser of a brand *W-Fashion*, advertiser here examines the features *Travel is a pleasure* and *Starbucks* and decides to integrate a user associated with a *Starbucks* item to the post of current round. For that, `b2biers` uses the *post concept* (U4) and *post coherence* (U5) units to search for suitable (satisfying the O1-criteria) *Starbucks* users. Three indicative such users are depicted in pictures 1 to 3 (taken from Pinterest) and are relative to an *entrepreneur*, *adventure*, and *family* profile, respectively. Advertiser finally selects the *entrepreneur* user for the post of round since the most recent posts liked by that user align better (based on U4 and U5 results) with the most recent posts published in the other features of post; *W-Fashion*, *Travel is a pleasure*, and *Starbucks*. Last, note that `b2biers` by using the *feature similarity* (U1) unit could also repeat the mentioned process for the features *Nescafe* or *Lavazza* or *Illy* (not present in *non-eliminated* set) that are similar enough with *Starbucks* based on U1 results; advertiser is the one who specifies whether the unit U1 is used or not for the O1 aims to enlarge further the search process.

page audience so as the brand to get known to other users besides subscribers. To achieve that, advertiser selects from the PDE *of* `b2biers` *the operation* O2. Figure 7 *depicts such a scenario where* `b2biers` *seeks to form the post of current round (round* 5*) that contains* $k = 4$ *features; the first feature is* W-Fashion. *Indicatively, the posts* p_1 *and* p_2 *are compared, with* p_1 *having the features* $\{$W-Fashion$, f_2, f_5, f_8\}$ *and* p_2 *the features* $\{$W-Fashion$, f_1, f_2, f_5\}$. *Eventually, the post* p_2 *is chosen by* `b2biers` *for the round* 5 *since it is estimated that it can influence two* 1-hop *users* (v_6, v_7) *and one* 2-hop *user* (v_{10}) *compared to only* v_6 *and* v_7 *influenced by* p_1. *Note that this process is just an instance of* O2*; namely,* ***all*** *the* $k-1$ *features in non-eliminated set should be checked* ***before*** *the decision for the post of round is taken.*

O3. Subscribers Influential Engagement. The operation of *subscribers influential engagement* finds the k features that maximize the influence on S (the set of subscribers to brand page) by *concurrently* influence a *fixed but different* portion of subscribers that were influenced before. This operation considers that each subscriber has a specific *like threshold* to get influenced but that threshold is *not known.* The goal of the operation O3 is to apply a *reinforcement learning* method to learn as much as possible closer values to those thresholds over rounds since the problem unfolds in many rounds.

This operation relates with the *retaining* aspect of *loyalty marketing* where the goal is to retain the interest of subscribers by finding ways to engage with them. Our engagement proposal is based on forming a content that gains the *maximum attention* of some subscribers in *current* round and some other subscribers in *previous* rounds (by not letting at the same time *big intervals* where a subscriber does not like the posts of brand). The aim is over rounds to include more subscribers in the process by simultaneously keeping them motivated.

The *technical challenge* of O3 is to *learn* how to *totally engage* (over rounds) the *biggest* portion of S. There should be a method that effectively influences subscribers stayed inactive, yet the operation success depends on who of them will influence and on what time (which round) it does it during the engagement process. We stress that all the *units* can be used for the *effective* and *efficient* execution of this operation apart from the *post coherence* unit.

Case 3. *Advertiser of* W-Fashion *observes* now *that several subscribers do not like the published posts in the page of* W-Fashion *with* satisfactory frequency. *Namely, there are* inactive intervals *for a number of rounds for several subscribers and that* restricts *the propagation of posts to the network since the subscribers of a brand are the initial adopters of brand. To alleviate this problem, advertiser executes from the* PDE *of* `b2biers` *the operation* O3. *A good applicability scenario of* O3 *can be seen in the first* 4 *rounds of* Figure 7 *where each user get influenced* twice *over rounds. To achieve that,* `b2biers` *can make use of the* feature similarity *(*U1*) unit to e.g., find similar features to influence user* v_1 *in rounds* 1 *and* 3. *Another use is the exploitation of* U2 *and* U3 *units we mentioned in* Case 2 *that can find stronger connections among subscribers utilized for influence-estimation*

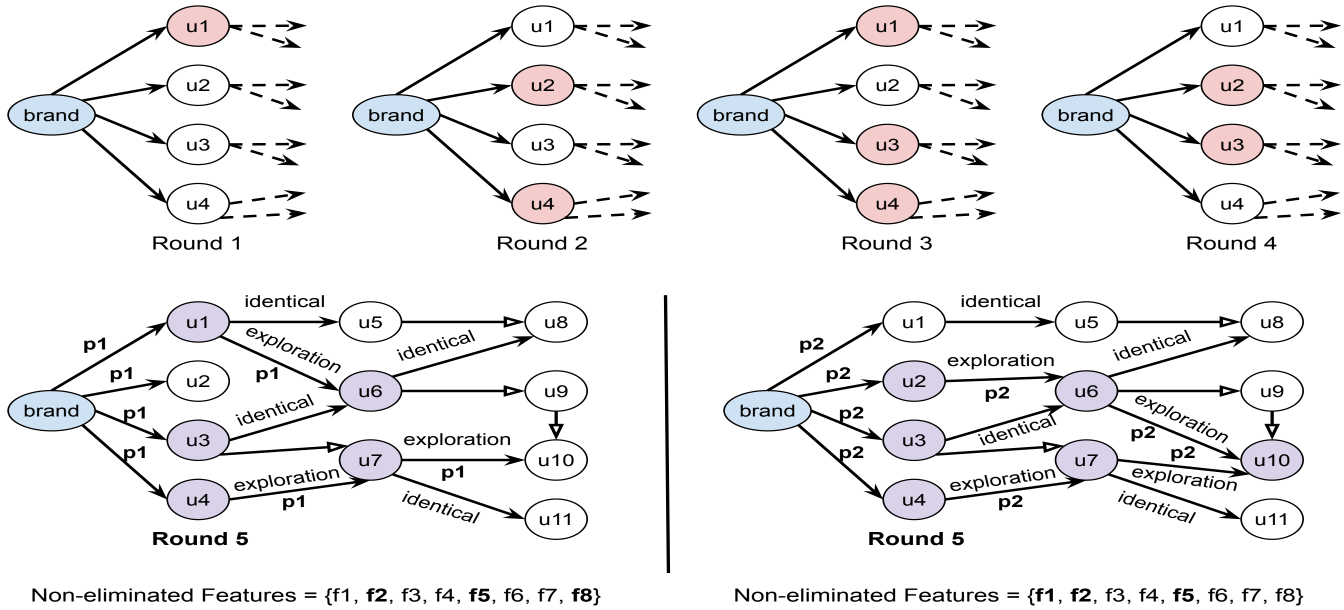


Figure 7: The brand *W-Fashion* publishes posts that *mostly* influence its subscribers; v_1 in round 1, v_2 and v_4 in round 2, v_1, v_3, and v_4 in round 3, and v_2 and v_3 in round 4. So, in round 5, advertiser of *W-Fashion* selects O2 of `b2biers` to form a post that can influence users beyond subscribers' level. For that, `b2biers` uses the *user identical content joins*; in short, *identical* (U2) and *user exploration content joins*; in short, *exploration* (U3) units to examine the influence potentials of candidate posts; here, p_1 and p_2. Note that only *exploration joins* affected by the selection of post, since *identical joins* have been formed based on the recent post *likes* of users; also, a connection among users without a label (*identical* or *exploration*) means that these users are friends in the respective social network (for simplicity, we show only one-directional connections). Suppose that a user *is expected* to get influenced if the user is associated with *more-than-one* influence connections and that the post propagates from a user to another (when the former likes the post) through an *identical*, *exploration*, or *friendship* connection. The post-propagation over a *friendship* connection naturally happens in social networks whereas post-propagation over *identical* and *exploration* connections happens with the use of recommendation algorithms employed by social network companies; also, advertiser via *messaging* can convert an *identical*/*exploration* connection to a *friendship* one for their aims.

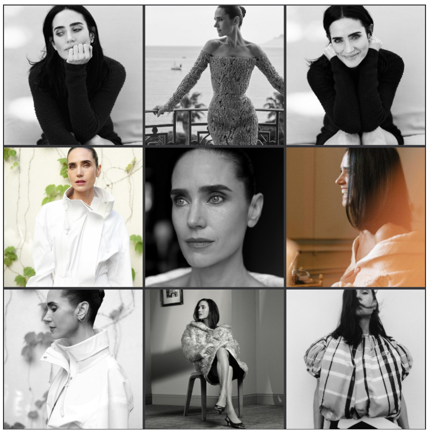
purposes by `b2biers`*; e.g., users v_3 and v_4 may be connected via an* identical *or* exploration *join in round* 3 *that enhances the influence chances for both of them. Last, the* post concept *(*U4*) unit can also be used; e.g.,* `b2biers` *finds the concept of the post that influenced v_2 in round* 2 *and utilizes this knowledge by proposing to advertiser a similar concept for the post of round* 4 *that can influence v_2 again.*

O4. Influential Post Diversity. The operation of *influential post diversity* forms *influential* and *diverse* k-size posts over rounds that are also *coherent* with previous posts to suit to the page audience. We emphasize that *diversity* and *coherence* are similar but different terms. The former relates with speaking about the same subject in a different way (e.g., under another concept, using similar but different features). The latter is already explained in our discussion for the *post coherence* unit. To clarify better the connection among *diversity* and *coherence*, consider the next example:

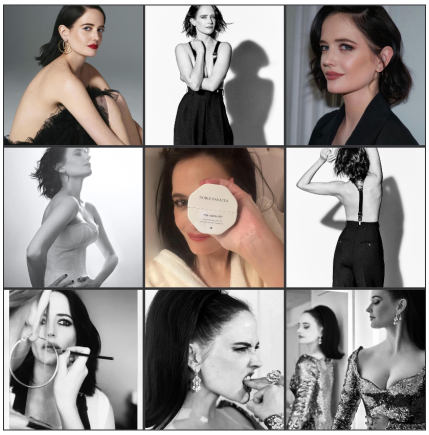
Example 5. *The posts* Fig. 4.3 *and* Fig. 5.3 *are diverse and coherent as separate aspects of mountain are shown in a natural way. The posts* Fig. 3.1 *and* Fig. 3.2 *are diverse but not coherent since* Yorgos *does not appear in the movie and* Emma *is completely different in real life than her appearance in "Poor Things". The posts* Fig. 3.2 *and* Fig. 3.4 *are not diverse but coherent because* Yorgos *has a very characteristic direction style. The posts* Fig. 5.1 *and* Fig. 5.2 *are neither diverse nor coherent as they both depict a public appearance of* Pom *in distinct outfits.*

Diversity of posts is important as it adds *naturalness* to brand page and makes its *exploration* (scrolling) interesting and not boring. E.g., if a brand page named "Life in Mountain" speaks about mountains, the posts Fig. 4.3 and Fig. 5.3 satisfy that purpose. So, post diversity can help a new visiting user to really like the brand page and may subscribe to it due to its *broad* and *connected* material.
The *technical challenge* of the operation O4 is to schedule a *policy* via which the k-size posts over rounds to be influential and at the same time diverse and coherent among them. Diversity could be captured via associating *entity tags* to features by categorizing them in a detailed depth; e.g., separate the *celebrity* features to *actors*, *singers*, *athletes*, etc., and then separate *actors* to *action*, *romance*, *comedy*, etc. The posts comparison based on their *entity tags* and *concept keywords* will define how much *diverse* and *coherent* they are, respectively. The goal is each post (of each round) to be maximum influential in the network and maximum diverse and coherent with a *specific* number of previous posts. The main units that can be used for this operation are the *feature similarity*, *post concept*, and *post coherence* units.

Case 4. *Advertiser of* W-Fashion *has noticed that the posts published in their page lack of* adequate diversity *and that creates a* repetitive experience *to users that hinders them from staying engaged with the brand* W-Fashion *(e.g., subscribers do not like often the posts; the posts reach only to few non-subscribers, etc.). To overcome this challenge,*



(a) Round 1 | Jennifer Connelly, elegance, class

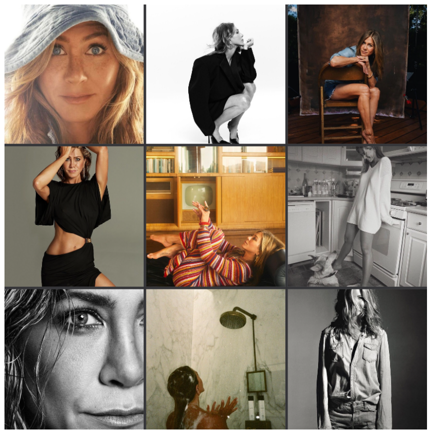


(b) Round 2 | Eva Green, beauty, style

(c) Round 3 | Charmaine Sheh, asian, charm

(d) Round 4 | Dior, Cindy Crawford, galerie

(e) Round 4 | Angelina Jolie, business, style



(f) Round 4 | Jennifer Aniston, casual, home

Figure 8: The posts published in the first 3 rounds of brand *W-Fashion* along with three proposed posts of `b2biers` to advertiser of *W-Fashion* to select from for the post of round 4. Each post comprises $k = 4$ features with the first feature being the *W-Fashion*; the other three features of posts are shown in the relative *caption* of each post where with capital-starting-letters we denote *brand* features (such as *Angelina Jolie*) and with all-small-letters we denote *abstract* features (like *elegance*) corresponding to a e.g., page that speaks about *elegant ideas and products*. The features mentioned for each post are *indicative* in order to highlight why each one of the proposed posts for round 4 is *diversified* to the posts published in prior three rounds. E.g., *Angelina Jolie* is also an actress of similar *age* and *style* with previous actresses but she is integrated to a *businesswoman-story* (including 8 other women) that differs from a *persona-story* that is built around 1 woman in the first three posts. Moreover, *Jennifer Aniston* is also an actress of similar *age* and *charm* with the actresses in the first 3 rounds, but she is *blonde* (not *brunette*) having more *light-appearance* characteristics and her post is relevant to more *casual* style, products, and environments than the posts of first 3 rounds that relate with more *luxury* and *classy* looks and outfits.

advertiser executes from the PDE *of* `b2biers` *the operation* O4. Figure 8 *depicts such a scenario where the posts in the first* 3 *rounds lack of* satisfactory diversity*; each of them shows a* brunette *actress of* similar-age *who can be described by* keywords *such as* elegance, class, beauty, style, and charm. *There are some differences that diversify these posts, such as* Charmaine Sheh *is* Asian *while the other two actresses are* Americans*, and* Eva Green *has a* baby-face *appearance compared to the more* mature-looks *of others, yet these differences can be considered* minor *to achieve* diversity. *So, suppose that in current round (round* 4*),* `b2biers` *proposes three posts to advertiser to select from for* O4 *goals, shown in* Fig. 8d, Fig. 8e*, and* Fig. 8f. `b2biers` *yields these results by using the* feature similarity *(*U1*),* post concept *(*U4*), and* post coherence *(*U5*) units to compare the post of round* 4 *with the posts of the prior three rounds.* E.g.*, for* Fig. 8d, Dior *relates with* elegance, class, beauty, style, and charm *but is a* company*, not a* person*;* Cindy Crawford *is of* similar-age *with the women in the first three posts, but she is a* model*, not an* actress*; and* most importantly*, a* galerie-story *relates but it is different from a* persona-story.

O5. Adaptive Subscription Maximization. The operation of *adaptive subscription maximization* selects k features to form posts (one post per round) so as to maximize the *cumulative subscription* over all rounds. Note that the respective operation for *cumulative influence* has already been implemented in [65] and constitutes the PDE component of `b2biers` we discuss in Section 2. The operation of *adaptive subscription maximization* can make drastic use

(a) Lucy Liu, asian, luxury **(b)** Robin Wright, casual, home

Figure 9: The Asian actress *Lucy Liu* and the American actress *Robin Wright*. The former represents similar things with *Charmaine Sheh* and with the respective concept of post in Fig. 8c; the latter shares similarities with *Jennifer Aniston* and with the respective concept of post in Fig. 8f. Above pictures used to justify our *applicability case*; they do not capture the *whole personality* of referred actresses.

of *user identical content joins* and *user exploration content joins* units to find *feasible targets* to gain their subscription; the *feature similarity* unit can also be used for an exhaustive search of *feasible subscription targets*.

The *technical challenge* of O5 is the deployment of an effective *reinforcement learning* technique that achieves a *repetitive influence* to *non-subscriber* users to gain their subscription. Several *repetitive influence* policies can apply; e.g., (i) if a user likes a *specific* number of times the posts of brand (independently of the content in them) then the user subscribes to brand page, (ii) if a user likes different aspects of posts' content related with a *specific* variety of user interests then the user turns to subscriber. We stress that O5 is more complex than our PDE problem in [65] and its successful implementation can lead to a second-PDE tailored *only* for *subscription* operations.

Case 5. *Advertiser of* W-Fashion *has* now *a different objective than prior cases; the focus now is on effectively* increasing *the* subscribers *of* W-Fashion *rather than creating* viral *posts for* influence *purposes. The reason is that advertiser noticed that while their previous published posts liked by a good amount of users, no* new *users* subscribe *to the brand page in a* satisfactory frequency. *So, assume that advertiser selects in this round the* indirect *(no* messaging *used) operation* O5 *from the* PDE *of* `b2biers` *to maximize the gaining of* new *subscribers to the page of* W-Fashion. *Suppose that* `b2biers` *uses the* U2 *and* U3 *units we discussed in* Case 2 *to find the targeted-for-gaining-their-subscription users* v_6 *and* v_7 *depicted in* Figure 7*;* v_6 *pertains to the aforementioned* repetitive influence *policy* (i) *and* v_7 *to policy* (ii) *with each user to* subscribe *to* W-Fashion *after influenced* twice *(from two different posts). Assume also that the posts* p_1 *and* p_2 *of* Figure 7 *map to the posts of two consecutive rounds (*p_1 *published before* p_2*) proposed by* `b2biers` *to advertiser for* O5 *aims, with* p_1 *being the post in* Fig. 8c *and* p_2 *the post in* Fig. 8f. *The user* v_6 *likes both posts because* v_6 *has a* general *preference on* aesthetics *around* fashion *and* celebrities *whereas the user* v_7 *likes* p_1 *due to the* specific *preference of user on* asian *and* luxury *lifestyles and* p_2 *due to* specific *preferences on* casual *clothing and* home-interior *environments. Via this process, both users* v_6 *and* v_7 *subscribed to the page of* W-Fashion*, even though they pertain to different* repetitive influence *policies. Finally, in case that* v_6 *and* v_7 *should be influenced* 4 *times (instead of* 2*) to gain their subscription,* `b2biers` *could additionally make use of* feature similarity *(*U1*) unit to form a similar post to* Fig. 8c *by replacing* Charmaine Sheh *with* Lucy Liu *and published by advertiser in the next-round-after-*p_2*, and a similar post to* Fig. 8f *by replacing* Jennifer Aniston *with* Robin Wright *and published by advertiser in the next-consecutive-round. This decision derived from the* U1 *results that show that* Charmaine Sheh *is similar to* Lucy Liu *and* Jennifer Aniston *is similar to* Robin Wright *(check* Figure 9 *for extra reference).*

O6. Gaining Subscribers by Messaging. The operation of *gaining subscribers by messaging* applies various *policies* under which *messages* are sent by the advertiser of brand to a *specific* number of *non-subscriber* users acting like an *invitation* to users to subscribe to brand page. A user can be notified (via a message) **only once**. We discuss three such representative messaging policies expressed as (k, m)-query, k-query, and m-query (all the *units* of `b2biers` can be applied in a variety of different combinations).

O6.1. (k, m)-query. In each round, advertiser searches what k features (forming one or more posts) to publish in brand page and which m users to notify via messaging of those k features so as to maximize the *cumulative subscription gain* over all rounds; *subscription gain* is defined as a weighted sum of the aggregate preference of m users to k features. This problem is first studied by us in [67].

The *applicability* of (k, m)-query is useful when advertiser does not have concrete ideas what to publish in brand page and so wants to explore which content (k features) would be promising. After running several (k, m)-queries then advertiser may focus on specific aspects of content that seem more suitable to page audience.

The *technical challenge* of O6.1 is the big number of k-feature-combinations (over *non-eliminated* features) and the huge number of m users present in social networks. Also, this problem can be solved *beforehand* over all rounds (*static* user preferences as in [67]) or *adaptively* over rounds (*dynamic* user preferences); in contrast to [67], `b2biers` requests to address O6.1 in bigger social networks (higher number of users and features) over *dynamic* user preferences.

Case 6. *Advertiser of* W-Fashion *applies* now *a more* aggressive *policy than* Case 5 *for gaining subscribers that relies on* direct messaging*; advertiser directly messages a user to subscribe to their brand page. For that, advertiser has three operations to select from the* PDE *of* `b2biers`*; here, we discuss the operation* O6.1 *relative to* (k, m)*-query. As said,* (k, m)*-queries used for* exploration purposes *so as advertiser to find what kind of content works well (gets several* likes*) for their brand page by gaining new subscribers at the same time. Suppose that (for* $k = 4$*) advertiser publishes, in a series of four consecutive rounds, the posts* Fig. 8d, Fig. 8e, Fig. 8f, and Fig. 8a*, respectively; namely,* `b2biers` *executed four times the operation* O6.1 *which was selected by advertiser to form the post of each round. Assume that results showed that the posts* Fig. 8f *and* Fig. 8a *play better (are more appealing to users) than the posts* Fig. 8d *and*

Fig. 8e*, and also the post* Fig. 8a *is more influential than the post* Fig. 8f*. Guided by these results, advertiser decides from now on, to publish posts that relate with* persona-stories *built around a* celebrity *woman (e.g.,* actress*,* model*, etc.) in an* elegant*,* stylish*, and* luxury *way; e.g., the next two posts of advertiser could be the posts* Fig. 8b *and* Fig. 8c*.*

O6.2. k-query. In each round, advertiser looks for the m users who are most interested to *predefined* k features. The difference from previous query is that now the k features are *already* selected by the advertiser. In many realistic scenarios, advertiser already knows with which other popular k features of network their brand content looks like. So, advertiser opts to notify m users of such k features.
The *technical challenge* of O6.2 is the need for its *real-time* applicability. This happens since the smaller search space (compared to O6.1) enables the deployment of *exact* solutions instead of also *approximate* ones that affect the solution accuracy. Further, k-queries are more often to be found in real world than (k, m)-queries, so many advertisers may *concurrently* apply k-queries to `b2biers`, and so their *real-time* response is crucial. Lieu in, the chances for accuracy loss and slower execution of (k, m)-queries are permissible due to the *content exploration* targets of those queries.
A *first approach* to efficiently address k-queries has already been implemented by our research team and submitted for publication to a premier computer science journal. We deployed several algorithms to solve the k-query problem for different and realistic query types over many users having dynamic preferences. Our best algorithms *actually* solve the problem in *real-time* by achieving significant superiority over its baselines for uniform and non-uniform queries. Nevertheless, similar to O6.1, the realistic nature of `b2biers` requests to address O6.2 in much bigger social networks.

CASE 7. *Assume that this* applicability case *takes place after previous* Case 6*. Moreover, suppose that the most posts in the page of* Jennifer Connelly *look* quite similar *to the post of* Fig. 8a *and the same holds for the pages of* Eva Green *and* Charmaine Sheh *and their published posts in regards to the posts* Fig. 8b *and* Fig. 8c*, respectively. Following the outcome of* Case 6*, advertiser of* W-Fashion *knows* now *that the content published in the page of* W-Fashion *is* similar enough *to the content published in pages* Jennifer Connelly*,* Eva Green*, and* Charmaine Sheh*, and for that advertiser opts to execute the operation* O6.2 *from the* PDE *of* `b2biers` *to gain a good amount of new subscribers in current round. In other words, advertiser opts to execute a k-query to* message *the* found *m users for the $k = 4$ features;* W-Fashion*,* Jennifer Connelly*,* Eva Green*, and* Charmaine Sheh*. Such a* message *would be similar to this one: "Hi user (username), if you like* Woman Fashion *and pages such as* Jennifer Connelly*,* Eva Green*, and* Charmaine Sheh*, you can visit my page* W-Fashion *and subscribe if you like it.".*

O6.3. m-query. In each round, advertiser asks for the k features that are more possible to stimulate the interest of given m users. Although this query seems to be inverse to k-query, it is actually more complicated than both previous queries. The reason is that the *coverage* of m users (stimulation on them) can be done in a *general* manner (based on aggregated user preferences) and in a *timely* manner (based on most recent user preferences that can depict a post concept similar to the one expressed by k features).
The *technical challenge* of the operation O6.3 is the fast and accurate implementation of a *combined coverage* method (*general* and *timely* manner oriented). Note that in practice, the logic of coverage pertains to more than one posts per round, and that makes the problem of m-query even more interesting and technically challenging. Cover a user entails collect the x_1 things that user liked the last x_2 days and convert them to posts where each one has a concept expressed by a subset of x_1, and do that in a recent-time priority way over x_2.
This coverage process is time-consuming and would prohibitively increase the overhead for the (k, m)-query problem. Still, note that this coverage process can also apply to k-query problem, but the number of users (searched) is much higher than the number of features (given) in social networks, and that would make the k-query coverage cumbersome in real-time compared to m-query problem. In the latter, the users are given and the features are sought, and that enables more sophisticated techniques (like coverage) to take place for gaining the subscription of users.

CASE 8. *Suppose that it is very important for advertiser of* W-Fashion *to gain the subscription of users v_8, v_9, v_{10}, and v_{11} shown in* Figure 7*. Yet, advertiser was not able to gain their subscription by using the previous operations* O5*,* O6.1*, and* O6.2*, due to the more distant connectivity of such users with the brand* W-Fashion *as can be seen in* Figure 7*. So, advertiser opts to execute the operation* O6.3 *from the* PDE *of* `b2biers` *to gain the subscription of mentioned users; namely, advertiser in current round applies a m-query to find the $k = 4$ features that can stimulate the interest of specific $m = 4$ users v_8, v_9, v_{10}, and v_{11}. Assume these users are* women *who generally like* fashion*-oriented posts and recently liked,* first*,* business*-oriented posts,* then*,* travel*-oriented posts, and* last*,* exhibition*-oriented posts.* `b2biers` *by utilizing this knowledge, proposes to advertiser to publish in current round the posts* Fig. 8e*,* Fig. 6*, and* Fig. 8d*. For simplicity, consider that each one of such posts is characterized by one* main feature*; that is,* business*,* travel*, and* galerie*, respectively. So, advertiser first publishes these three posts in the page of* W-Fashion *and then notifies each one of the aforementioned $m = 4$ users with a* message *similar to that: "Hi user (username), if you like* Woman Fashion *and content about* business*,* travel*, and* exhibitions*, you can visit my page* W-Fashion *and subscribe if you like it.". Finally, note that there is a* time-interpretation *relative to the* coverage *of m users. Namely, each visiting user to* W-Fashion*, first see the post* Fig. 8e *that covers their daily business life, then they see the post* Fig. 6 *that covers their need for travel, and last they see the post* Fig. 8d *that covers their desire for visiting interesting places in their travel destinations. This* timely-based *coverage, satisfied by* `b2biers` *for the* O6.3 *purposes, significantly enhances the chances of referred m users to subscribe to the page of* W-Fashion*.*

4. RELATED WORK

4.1 Influence Maximization

The *classic* Influence Maximization (IM) problem seeks for the k users who can maximize the influence of a *given* post in a social network. Particularly, Kempe et al. [35] were the first that formulated the problem based on the Independent Cascade (IC) and Linear Threshold (LT) propagation models, proved its NP-hardness, and proposed a greedy algorithm with approximation guarantees. Subsequent works investigated efficiency and scalability questions, either with heuristics [17; 19; 22] or preserving an approximation guarantee [43; 28; 21; 9; 64]. IM is a very popular problem studied extensively the last two decades due to its *viral marketing* effect [24] that drastically spreads a new product in a network. The work [47] presents a detailed survey on IM and discusses several of its variants. Instead of maximizing the influence of a *fixed* post by searching for influential initial adopters (classic IM), `b2biers` depends on the idea of CAIM problem [33] that given initial adopters (brand's subscribers), it searches for k influential features (social network pages) to form the content of a *tunable* post to maximize its influence in a network. Other works also utilize *content* to achieve influence in a social network [2; 4; 3; 15; 48; 36], yet differ from CAIM [33] in terms of targets, content type, initial adopters, or propagation models.

The *adaptive* Influence Maximization (AIM) problem maximizes the *cumulative influence* of a *fixed* post in a social network over *many rounds* [28] by selecting influential adopters based on *network feedback* (the users set that liked the round post). Some works study this problem under *known* network parameters [27; 20; 78; 62; 32] while others solve it by concurrently learning *latent* parameters [42; 18; 70; 72; 73]. Differently from all mentioned adaptive IM works, the PDE of `b2biers` [65] (discussed in Section 2) solves AIM in many rounds by searching for content (features) to form *tunable* posts (one post per round). Also, the propagation model of PDE (CATRID in [65]) depends on node (user) activation probabilities and not on edge (user-user connection) activation probabilities leveraged in prior AIM works.

4.2 Components of b2biers

We remind that by the term *components* we mention to the 5 *units* and 8 *operations* we proposed in Section 3. Here, we discuss how each one of such *components* is related with existing literature and why they constitute *novel problems*.

U1. Feature Similarity. As mentioned, we have recently published our first work [66] on *feature similarity*. This work finds how similar are two given communities (brands or features or pages in this article). Yet, we stress that feature similarity component remains a *novel proposal* since the final aim is to implement an *all-pair* (all pairs of features) and *dynamic* (preferences of user change over time) join operator on feature similarity that will significantly extend our *single-pair* join operator in [66]. In regards to the general literature, feature similarity constitutes a new and alternative variant of classic ϵ-join operator [6; 49; 34] that finds all pairs of points within ϵ distance to each other among two d-dimensional datasets. The three differences of feature similarity join as proposed in [66] and classic ϵ-join is that the former: (i) relies on finding one-to-one user pairs instead of all user pairs among datasets, (ii) applies the ϵ condition per dimension and not over all dimensions in an aggregated way as e.g., Euclidean distance of ϵ-join, and (iii) uses a meaningful value for ϵ avoiding issues relative to the choice of a proper ϵ value in regards to the selectivity of the join.

U2. User Identical Content Joins. We remind that this unit finds for each user v a number of m users who like at least k identical posts with v over a given time period. This problem can be modeled as a *variant* of the popular kNN *join* problem [7; 8] that finds the k nearest items for each user and applies to a lot of cases in real world. It pertains to *exact* and *approximate* solutions, yet our proposed join variant only relates with exact solutions. Authors in [68] present a recent and complete survey on exact kNN joins. In particular, the problem we propose is a *self-user* kNN join that instead of being based on a distance metric (such as Euclidean distance) it depends on liking identical content. All the existing literature on kNN joins [68] utilizes techniques based on a distance metric, while the variant join we propose (that relies on a different evaluation metric) can be solved in a much faster way by issuing a *reinforcement learning* approach as explained for U2 in Section 3. During implementation of proposed unit, we intend to compare our new join (U2) with the most relevant works on exact, real-time, and dynamic kNN joins such as the works [74; 69].

U3. User Exploration Content Joins. The related work of U3 is the same with the one of U2. Additionally, U3 can also be compared with the top-k join works in [13; 14]. The reason is that such works find the top-k pairs of the objects (among two datasets) that are similar to each other in one subspace and dissimilar in another subspace. Although this kind of joins capture the targets of U3 as explained in Section 3, the *novelty* of U3 relies on two factors: (i) U3 is a kNN join and not a top-k join, (ii) U3 utilizes the ϵ-based absolute difference condition of [66] and not the Euclidean distance or any other generic function leveraged in [13; 14].

U4. Post Concept. This unit that *given* a post it tries to find the k keywords that describe the post (capturing its *concept*) relates with the *contextual search* literature [30; 5; 38; 37]. Contextual search augments the search query (or background if query not given) of a user with the recent search history of user (that defines the user search context) so as to find more relevant results to the user needs. However, there are three main differences of U4 in regards to contextual search. First, searching in search engines differs from browsing in social networks, so the context in such two cases carries different semantics; in the former, the users have a clear focus to find something specific, while in the latter, the users can like whatever post they find attractive during their browsing. Second, U4 handles only a small portion of known information (mentioned in Section 3 as *ground truth* for brands), while contextual search depends on a large portion of known information. Third, U4 pertains to the processing of many users (as applies to a social network environment) and not only to a single user as happens in contextual search. So, U4 could utilize common points to contextual search, but it is a *clearly different* problem than searching with context and its variants [30; 5; 38; 37].

U5. Post Coherence. The unit U5 that examines whether some *given* k features can form a *coherent* post is correlated with works on *related item recommendation* [76], *online learning to rank* [46; 44], and mainly with works on *topic coherence* [55; 53; 57; 59]. The first approach (related item recommendation) solves a similar problem to U1. Namely, the idea is that if two features found similar then they are coherent too, so U1 could be utilized to solve the U5 problem. Nevertheless, that would be a restricted approach since similarity of features handles the subscriber sets of a feature pair and not the posts of features in a continuous way as U5 does it (see Section 3). In other words, U5 avoids cases where two features are generally similar but at the current point of time they are involved in posts that are not coherent; such cases that are frequent in real world cannot be identified by U1. The second approach (online learning to rank) addresses the problem where the results provided by a search engine should be relative (coherent) among them. Yet, the found relevant results are the solution to the problem while in U5 we have a given input that we want to check its coherence at the beginning of the problem. Also, as earlier mentioned, search engines and social networks are two completely different evaluation environments. Finally, the third and closest approach to U5 refers to the topic coherence problem; given a set of topics derived from a topic model, topic coherence evaluates how coherent are the found topics. The crucial difference of U5 with the topic coherence problem is that U5 aims to evaluate whether a set of *features* are coherent instead of a set of *topics*. As we analytically explain in our prior work [33], features and topics substantially differ. The former are social network pages that can be related with a lot of different things (each entity that has a social network page is considered a brand as it has a specific number of subscribers), while the latter are concrete terms operating as general descriptors of the relative content (document, article, etc.). To give a short example, the word *comedy* is a topic and every distinct entity that relates with comedy (among *actors*, *celebrities*, *movies*, and so on) is a separate feature. Therefore, U5 constitutes a *novel problem* that still could benefit from U1 and topic coherence works. As a side note, it would be interesting and challenging to investigate the connection of U5 with *storytelling* works [61; 16]. Storytelling is the technique that conveys visual information based on data analysis in an easily understandable way. Hence, it could contribute to the coherence of the features for the content-tuning purposes of U5.

O1. Influential User Post Integration. This operation, which intends to find a suitable influential user to participate (with one of their existing pictures) in the post of current round, relates with works on *influencer marketing* [26; 11; 10; 40] and *user-generated content advertising* [39; 23; 52]. The former search a set of influencers to advertise the products of a brand; the difference with classic influence maximization problem is that in influencer marketing the found influencers create a post depicting themselves using the brand product instead of just sharing a post created by the brand advertiser. The latter mention to the scenarios where simple users of the platform (social networks, e-commerce, websites, etc.) create any-type of personal content (text, video, picture, etc.) that advertiser could exploit in case that content relates with the brand that advertiser wants to promote. In particular, advertiser can either use that content along with its own professional content prepared for the brand, or advertiser can contact with the creators of generated content and together tune the final content for the promotion of brand. The operation O1 we propose relates with the case that advertiser uses a part of (and not whole of it) user-generated content to tune a *collage post* that promotes the relative brand; O1 seeks to integrate to a collage post a single piece of user-generated content of an influential user that is part of an influencer marketing solution. We claim that O1 is a *novel problem* since there is no work that computationally (in an algorithmic way) investigates such a co-creation for the content-tuning of a collage post as we discussed it in Section 3.

O2. Influential Paths Extension. The purpose of operation O2 is to increase over rounds the number of *connection paths* (connecting influenced users to brand's subscribers) on which it can rely for further *influence* and potential *subscription*. So, O2 relates with works [60; 31; 40] that *adaptively* find influential users under respective network knowledge over rounds. However, we stress that O2 achieves that purpose by searching for influential content (influential features) instead of seeking for influential initial adopters (seeds) as prior works do, and so O2 constitutes a *novel problem*. Besides that, we note that O2 also contributes to the *subscription* aspect that is not covered by previous adaptive works.

O3. Subscribers Influential Engagement. The operation O3, which tries to engage the maximum number of subscribers over rounds by not letting big intervals where a subscriber does not like a post of advertised brand, could be related with *personalized* [29; 48; 56; 40] and *uniform* [45] influence maximization works. The former maximize the influence on *targeted users* based on their preferences and the request of respective query. The latter seek for seeds who can influence a maximum number of users from as many *different communities* as possible. The objective of O3 relates with mentioned works since O3 personalizes the influence target to the brand's subscribers and it tries to achieve that target in a uniform way by engaging additional and different subscribers as rounds evolve. Nevertheless, the objective of O3 relies on feature-content-tuning and not on influential-users-seeking as happens in prior works. Still, the objective of O3 has more parameters to consider during implementation. For these reasons, the operation O3 is a *novel problem*.

O4. Influential Post Diversity. The goal of operation O4 is to find features for forming *influential*, *diverse*, and *coherent* posts over rounds. In most cases, diversity and coherence (or relevance) are studied combined in the literature that addresses the problem of diversified search results. Diversity of results has been studied in various domains such as databases, web search, information retrieval, and so on; some indicative works are presented here [25; 1; 58]. The operation O4 is mainly related to *diversified influence maximization* [63; 12; 45] and in a supplementary way to *diversified online learning to rank* [80; 44]. In regards to the latter, we already explained previously that social networks (our case) and search engines (online learning to rank) share some similarities but represent different evaluation environments. Also, O4 wants to achieve its purpose for every round and not spending several rounds till that can happen (as in online learning to rank works). The closest counterpart to O4 targets is the mentioned diversified influence maximization works, yet as already stated, such studies search for influential users to achieve their goals whereas O4 seeks for influential features to tune a suitable content to achieve its goals. Besides, O4 applies under multi-round settings in contrast to a single round taking place in [63; 12; 45]. These factors justify the *novelty* of discussed problem in O4.

O5. Adaptive Subscription Maximization. The execution of operation O5 pertains to *indirect* (more natural; without messaging) gained subscription that relies on the idea of *repetitive influence* to *non-subscribers*. Relative works to O5 lie in the *personalized influence maximization* literature [29; 50; 41; 48; 56; 40] that maximizes in various ways the influence on *targeted* users based on their preferences. Yet, O5 searches for appealing features instead of influential seeds to achieve its aims, and that differentiates it from prior works. Moreover, the objective of O5 is the *repetitive* influence to targeted users and not an *one-off* influence scheduled in previous works. So, O5 may benefit from some techniques deployed in mentioned works but significantly differs from them, and that makes it a *novel problem.*

O6.1. (k, m)**-query.** As mentioned, we recently introduced in [67] the problem of *gaining subscribers using content*, which pertains to an advertiser who manages the social network page of a brand that wants to increase the subscribers to its page. By using a (k, m)-query we search in each round both for k features (content) and m users to *send messages* containing such k features to gain the subscription of m users. The goal is to maximize the *cumulative subscription* over all rounds. However, our first approach on (k, m)-query was based on small networks (users and features) and static user preferences. Therefore, the (k, m)-query problem (operation O6.1) remains a *novel problem* since more sophisticated techniques need to be developed in order to more practically address the real-world settings of O6.1.

O6.2. k**-query.** The operation O6.2 is a *variant* of O6.1 and notifies m non-subscriber users for *predefined* k features, as in many real-world scenarios the k features are already known to the brand advertiser. Relative research to O6.2 that stimulates the interest of users based on keywords is present to literature [77; 71; 54; 51]. Yet, all mentioned works *search* for the proper keywords to achieve their purposes whereas the execution of O6.2 relies on *given* keywords (features). Also, the nature of utilized keywords (in their case) and features (in our case) differs as we also stated earlier for other components of `b2biers`. These reasons verify the *novelty* of presented problem in O6.2.

O6.3. m**-query.** The operation O6.3 has the *inverse* objective to O6.2; given m users it looks for k features to notify them. Since the number of m users is *limited* per round, the logic of O6.3 is to stimulate their interest in a more realistic and advanced way (by the notion of *coverage* as explained in Section 3) than both previous queries; we also discuss in Section 3 why coverage is not suitable for O6.1 and O6.2 problems. Coverage could leverage some *personalized influence maximization* works we mentioned for O5 but coverage could also benefit from *active friending* works in social networks [75; 79]. Active friending is a recommendation strategy that guides the interested user to systematically approach their *specific* friending targets. Referred works maximize the probability of friending targets to accept the invitation of interested user (become social friends with user). Although gaining the friendship of users differs from gaining their subscription (that is the goal of O6.3), active friending could contribute in a complementary way to the subscription purpose. To conclude, O6.3 is a *novel* and conceptually more complex problem than O6.1 and O6.2.

5. ADVERTISING TECHNIQUES IN THE SOCIAL NETWORK INDUSTRY

In this section, we show the advertising techniques that take place in the social network industry to clarify that our work *does not overlap* with such techniques. Yet, we stress that `b2biers` can also apply the presented techniques additionally to its goals; namely, `b2biers` offers *new advertising services* without cancelling or affecting the advertising techniques of industry. Specifically, we discuss advertising techniques that apply to the most popular social networks, such as Facebook[4], YouTube[5], and Instagram[6]. Since there is no literature that shows the actual internal advertising mechanisms of such companies, we found how advertising takes place in them by Neil Patel[7], one of the top 10 marketers according to Forbes[8]. All the techniques and their details we present here derived from the YouTube playlists of Neil Patel named *Facebook Unlocked*[9], *YouTube Unlocked*[10], and *Instagram Unlocked*[11]. Last, as most of the techniques are *common* among Facebook, YouTube, and Instagram, we avoid to discuss *repetitive* material over the following sections.

5.1 Facebook

Offer. An *offer* gives advertisers a chance to connect with prospective users. The offer should be easy to understand and having a compelling picture to stand out. Each offer has an expiration date.

Business Manager. The *business manager* is a platform that allows advertisers to manage all their brand (client) pages; share content among pages, select security settings, synchronize payments, etc.

Content Types. There are various *content types* that advertisers may publish in their brand pages:

- **How-To Posts.** They create questions and responses that yield engagement; a step-by-step process is used.
- **Video Tutorials.** They depict the visual expression of prior posts and so they can be more memorable.
- **Industry-specific Stats.** Such information relies on data to describe the marketplace production and consumption. Stats used in a way that are true but also show an unexpected result that attracts user attention.
- **Industry News.** Users like to discuss for such industry news (e.g., a new algorithm published by Google).
- **Case Studies.** Similar to *Industry-specific Stats* but they have a more detailed and general focus.
- **Checklists.** A list of things one have to consistently and gradually do to achieve a specific goal.
- **Weekly Roundups.** A collection of tools, strategies, and statistics discussed by a group of experts.
- **Instructional Guides.** Short articles that usually leverage visual elements to express their information.
- **Podcasts.** Podcasts are an audio (conversational) medium and they optionally contain also a video component. They consist of episodes that build intimate relationships with audience.

[4]`https://facebook.com/`
[5]`https://youtube.com/`
[6]`https://instagram.com/`
[7]`https://neilpatel.com/`
[8]`https://forbes.com/`
[9]`https://www.youtube.com/watch?v=LInnulL_6is&list=PLJR61fXkAx10aFOfkKsQCX9OHFiqwK72S`
[10]`https://youtube.com/watch?v=BJhTePXFvGo\&list=PLJR61fXkAx13HymYam75l8XRyayA3ail4`
[11]`https://youtube.com/watch?v=eJ2NNy1F6y4\&list=PLJR61fXkAx13W2yl_3VdXqspy4-5DCKKm`

- **Quotes.** Short sentences usually phrased by popular people and used to describe different situations.
- **Vlogs.** Videos depicting the daily life of people known as vloggers (derived from video bloggers). They also comprise episodes; often 1 vlog uploaded per day.
- **Webinars.** Live videos shared with users in real time.
- **Evergreen Content.** It mentions to a post that marked to be appeared as the first post in the brand page; it continues to be relevant to users over time.
- **Testimonials.** Posts mentioning to positive feedback of users to brand products that such users consumed.

As shown, there are several *content types* that can have success. However, we stress that *no one size fits all*. E.g., some users like video posts, while other users prefer posts that mainly comprise text. Advertisers should try different things and stay in the long-run with the content types that clearly yield more engagement than others.

Redistribution. Advertisers should distribute their material on multiple platforms (other social networks, websites) to create several touch-points with users. Content distribution can be in different formats (e.g., extract quotes from a video) and successful content can be redistributed (even on same platform) with a different format.

Word of Mouth. Advertisers notify their familiars (friends, family, neighbors, etc.) about their brand pages.

Cooperation. Advertisers partner with other advertisers who manage other pages (of similar topic) having similar number of subscribers and likes per post. This boosts post views for both parties while providing good content they did not need to create since each advertiser creates content for their own page.

Reports. There are several *reports* that advertisers can advise to observe statistics relative to the posts they publish. Based on such reports, they can identify high-performing content, compare individual posts, learn the characteristics of the users that react to posts, and so on.

Paid Advertising. Besides publishing posts to promote their pages, advertisers can also pay the platform (here, Facebook) to advertise a post that they select (already published in their page). In more detail, advertisers can select parameters such as *location*, *age*, *gender*, and *interests* of users for their post so as the advertisement algorithm of Facebook to make it targeted to a proper audience. More advanced features relative to targeting, allow advertisers to create a *custom audience* either by uploading a list of users (e.g., good brand customers) or by locating a page on which Facebook via automatic tagging to the subscribers of that page, can find similar users. Paid advertising on Facebook also relates with the following:

- **Budget.** Via a *trial and error* process, advertisers should learn to spend wisely their budget that allocate for advertising their posts. Facebook uses an algorithm that shows more often advertisements to users that yield engagement (*likes*, *comments*, *shares*). So, it does not mean that more money spent on advertising can yield more engagement. What matters is advertisers to learn what posts work for them and also to have a *high-quality* content to advertise. These things will help them to spend their budget more effectively.
- **Types.** (i) *picture*: pictures having more than 20% text may experience reduced delivery, (ii) *video*: the shorter the video the better, (iii) *instant experience*: it is a full-screen, post-click experience where users can swipe through carousels, complete a form, access the items easily, and discover lifestyle pictures with tagged products, and (iv) *collection*: it involves a cover picture or video followed by several pictures of the product; a click to a *collection* leads to *instant experience*.
- **Tools.** (i) *quality ranking*: a metric that compares the post selected by advertisers to other posts (of other advertisers) competing for the same audience, (ii) *engagement rate ranking*: same as previous but measuring expected engagement rate, (iii) *conversion rate ranking*: same as previous but measuring expected conversion rate (e.g., if a user likes a post and then subscribes to brand page in which that post is published), (iv) *ads manager*: similar to aforementioned *business manager* but specialized on advertisements, and (v) *ad library*: a library that contains advertisements that have success, so it can give ideas to advertisers.

5.2 YouTube

Video Content Types. Latest Trend | Challenges | Social Experiments | Comparisons | Q & A | Tours & Walkthroughs | Gaming Walkthroughs | How To Guides and Tutorials | Vlogs | Product Reviews | Lives & Webinars | Essays | Favorite/Best Of | Unboxing | DIY | Educational | Celebrity | Comedy | Travel & Lifestyle | Animal.

Keywords Research. Advertisers use *search engine optimization* (SEO) tools like the popular Google Trends (`https://trends.google.com/trends/`) and Ubersuggest (`https://neilpatel.com/ubersuggest/`) to find and expand keywords so as to create videos guided by such keywords. This helps the YouTube algorithm to apply a more effective ranking for their videos and so more YouTube users see them.

General Actions. Add download and competitor links | Use clever thumbnails | Use thank-you-for-watching comments | Ask to subscribe | Use a storyboard and automatic captions in the videos | Comment on competitor videos.

5.3 Instagram

Content Types. Stories | Giveaways | Tutorials | Quotes | Open-ended Questions | Before & After | Behind-the-Scenes | Interviews | Trends | User-generated Content | Influencers-advertising Content.

Influencer Marketing. Instagram is the social network of *influencers*; people having a lot of subscribers in their pages. So, many advertisers pay influencers to advertise their products; the relative posts uploaded by influencers are called *sponsored posts*.

User-generated Content. Instagram is the most popular social network for using pictures since many users upload every day several pictures depicting themselves. So, advertisers can identify relevant *user-generated content* to include it in their posts.

Giveaways. They are a type of contest often managed by *influencers* in behalf of advertisers where engagement messages are used such as: (i) like and share to win, (ii) follow our brand to win, (iii) tag a friend to win, and (iv) upload your picture of advertised product to win.

6. NOVELTY OF b2biers

We emphasize that there are three factors that justify the *novelty* of `b2biers`: (i) prior work, (ii) academic literature, and (iii) industry. We summarize them as follows:

Prior Work. As mentioned, the creation of `b2biers` depends on our prior work (PDE) [65]. Actually, the impressive performance of PDE in a variety of realistic case studies (as shown in [65]) motivated us to design the `b2biers` system. So, we declare that all the 5 *units* and 8 *operations* (13 components) presented in Section 3 are scheduled having the PDE in mind (discussed in Section 2) but we highlight that they are completely independent-to-PDE problems. The PDE just creates a *framework* to which such 13 separate problems can apply and altogether yield the `b2biers` system. Even for the *feature similarity* (U1), (k, m)-query (O6.1), and k-query (O6.2) components for which as stated we have already implemented preliminary versions, we need more sophisticated and advanced methods of them to be supported by `b2biers`. Specifically, our current version of *feature similarity* handles only a single pair of features with static user preferences, while we would need for `b2biers` an optimized all-pair feature similarity with dynamic user preferences. A dynamic solution is also necessary for the (k, m)-query problem, while both (k, m)-query and k-query problems request more efficient and scalable methods for bigger social networks. So, we stress that *all* the 13 components presented in Section 3 are *novel works* with their own *distinct* technical challenges, besides the fact that we scheduled them for PDE.

Academic Literature. We analytically explain in Section 4.2, how each one of the 13 components (presented in Section 3) we proposed in this article for the creation of `b2biers`, differs from the existing academic literature.

Industry. All the *services* (Section 3) provided by `b2biers` *do not exist* in major social networks such as Facebook, YouTube, and Instagram. This happens because all the industry techniques discussed in Section 5 are based on a *given content* whereas all the `b2biers` services *guide* the advertisers to *form a content* to advertise. A side benefit of that is that the industry techniques can additionally apply as a supplement to the `b2biers` services. Namely, `b2biers` not only does not cancel the industry techniques but also helps advertisers to have an *algorithmic*, *dynamic*, *broad*, and *consistent* way to form engaging (influence and subscription) content to promote their brand pages.

7. CONCLUSION

We presented our *design* for the deployment of `b2biers` system that can be the *first system* providing *open-source access* to *a variety of services* for maximizing *influence* and *subscription* in social networks based on *content.* We believe our contributions (5 *units* and 8 *operations*) are vital to the *open problem* of social network engaging advertising.

8. ACKNOWLEDGEMENTS

This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-RP-2021-020).

9. REFERENCES

[1] S. Abbar, S. Amer-Yahia, P. Indyk, and S. Mahabadi. Real-time recommendation of diverse related articles. In *WWW*, pages 1–12, 2013.

[2] S. Aral and D. Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639, 2011.

[3] N. Barbieri and F. Bonchi. Influence maximization with viral product design. In *SDM*, pages 55–63, 2014.

[4] N. Barbieri, F. Bonchi, and G. Manco. Topic-aware social influence propagation models. *KAIS*, 37(3):555–584, 2013.

[5] D. Billsus, D. M. Hilbert, and D. Maynes-Aminzade. Improving proactive information systems. In *IUI*, pages 159–166, 2005.

[6] C. Böhm, B. Braunmüller, F. Krebs, and H.-P. Kriegel. Epsilon grid order: An algorithm for the similarity join on massive high-dimensional data. *SIGMOD Rec.*, 30(2):379–388, 2001.

[7] C. Böhm and F. Krebs. Supporting kdd applications by the k-nearest neighbor join. In *DEXA*, pages 504–516, 2003.

[8] C. Böhm and F. Krebs. The k-nearest neighbour join: Turbo charging the kdd process. *KAIS*, 6(6):728–749, 2004.

[9] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier. Maximizing social influence in nearly optimal time. In *SODA*, pages 946–957, 2014.

[10] D. Brown and S. Fiorella. *Influence marketing: How to create, manage, and measure brand influencers in social media marketing.* Que Publishing, Indianapolis, 2013.

[11] D. Brown and N. Hayes. *Influencer marketing.* Routledge, London, 2008.

[12] A. Caliò, R. Interdonato, C. Pulice, and A. Tagarelli. Topology-driven diversity for targeted influence maximization with application to user engagement in social networks. *TKDE*, 30(12):2421–2434, 2018.

[13] M. A. Cheema, X. Lin, H. Wang, J. Wang, and W. Zhang. A unified approach for computing top-k pairs in multidimensional space. In *ICDE*, page 1031–1042, 2011.

[14] M. A. Cheema, X. Lin, H. Wang, J. Wang, and W. Zhang. A unified framework for answering k closest pairs queries and variants. *TKDE*, 26(11):2610–2624, 2014.

[15] S. Chen, J. Fan, G. Li, J. Feng, K.-l. Tan, and J. Tang. Online topic-aware influence maximization. *PVLDB*, 8(6):666–677, 2015.

[16] S. Chen, J. Li, G. Andrienko, N. Andrienko, Y. Wang, P. H. Nguyen, and C. Turkay. Supporting story synthesis: Bridging the gap between visual analytics and storytelling. *IEEE Transactions on Visualization and Computer Graphics*, 26(7):2499–2516, 2020.

[17] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, page 1029–1038, 2010.

[18] W. Chen, Y. Wang, Y. Yuan, and Q. Wang. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. *JMLR*, 17(1):1746–1778, 2016.

[19] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *ICDM*, pages 88–97, 2010.

[20] Y. Chen and A. Krause. Near-optimal batch mode active learning and adaptive submodular optimization. In *ICML*, pages 160–168, 2013.

[21] Y.-C. Chen, W.-C. Peng, and S.-Y. Lee. Efficient algorithms for influence maximization in social networks. *KAIS*, 33(3):577–601, 2012.

[22] S. Cheng, H. Shen, J. Huang, W. Chen, and X. Cheng. Imrank: Influence maximization via finding self-consistent ranking. In *SIGIR*, page 475–484, 2014.

[23] T. Daugherty, M. S. Eastin, and L. Bright. Exploring consumer motivations for creating user-generated content. *Journal of Interactive Advertising*, 8(2):16–25, 2008.

[24] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, page 57–66, 2001.

[25] M. Drosou and E. Pitoura. Search result diversification. *SIGMOD Rec.*, 39(1):41–47, 2010.

[26] P. Gillin. *The new influencers: A marketer's guide to the new social media*. Linden Publishing, Sanger, 2007.

[27] D. Golovin and A. Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *JAIR*, 42(1):427–486, 2011.

[28] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. A data-based approach to social influence maximization. *PVLDB*, 5(1):73–84, 2011.

[29] J. Guo, P. Zhang, C. Zhou, Y. Cao, and L. Guo. Personalized influence maximization on social networks. In *CIKM*, page 199–208, 2013.

[30] M. Henzinger, B.-W. Chang, B. Milch, and S. Brin. Query-free news search. In *WWW*, pages 1–10, 2003.

[31] T. Horel and Y. Singer. Scalable methods for adaptively seeding a social network. In *WWW*, page 441–451, 2015.

[32] K. Huang, J. Tang, K. Han, X. Xiao, W. Chen, A. Sun, X. Tang, and A. Lim. Efficient approximation algorithms for adaptive influence maximization. *VLDBJ*, 29(6):1385–1406, 2020.

[33] S. Ivanov, K. Theocharidis, M. Terrovitis, and P. Karras. Content recommendation for viral social influence. In *SIGIR*, page 565–574, 2017.

[34] D. V. Kalashnikov. Super-ego: fast multi-dimensional similarity join. *VLDBJ*, 22(4):561–585, 2013.

[35] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.

[36] A. Khurana, A. Logins, and P. Karras. Selecting influential features by a learnable content-aware linear threshold model. In *CIKM*, page 635–644, 2020.

[37] R. Kraft, C. C. Chang, F. Maghoul, and R. Kumar. Searching with context. In *WWW*, pages 477–486, 2006.

[38] R. Kraft, F. Maghoul, and C. C. Chang. Y!q: Contextual search at the point of inspiration. In *CIKM*, pages 816–823, 2005.

[39] S. Krishnamurthy and W. Dou. Note from special issue editors: Advertising with user-generated content: A framework and research agenda. *Journal of Interactive Advertising*, 8(2):1–4, 2008.

[40] P. Lagrée, O. Cappé, B. Cautis, and S. Maniu. Algorithms for online influencer marketing. *TKDD*, 13(1):1–31, 2019.

[41] J.-R. Lee and C.-W. Chung. A query approach for influence maximization on specific users in social networks. *TKDE*, 27(2):340–353, 2014.

[42] S. Lei, S. Maniu, L. Mo, R. Cheng, and P. Senellart. Online influence maximization. In *KDD*, pages 645–654, 2015.

[43] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *KDD*, page 420–429, 2007.

[44] C. Li, H. Feng, and M. d. Rijke. Cascading hybrid bandits: Online learning to rank for relevance and diversity. In *RecSys*, page 33–42, 2020.

[45] J. Li, T. Cai, K. Deng, X. Wang, T. Sellis, and F. Xia. Community-diversified influence maximization in social networks. *Information Systems*, 92:101522, 2020.

[46] S. Li, T. Lattimore, and C. Szepesvári. Online learning to rank with features. In *ICML*, pages 3856–3865, 2019.

[47] Y. Li, J. Fan, Y. Wang, and K.-L. Tan. Influence maximization on social graphs: A survey. *TKDE*, 30(10):1852–1872, 2018.

[48] Y. Li, D. Zhang, and K.-L. Tan. Real-time targeted influence maximization for online advertisements. *PVLDB*, 8(10):1070–1081, 2015.

[49] M. D. Lieberman, J. Sankaranarayanan, and H. Samet. A fast similarity join algorithm using graphics processing units. In *ICDE*, pages 1111–1120, 2008.

[50] Q. Liu, Z. Dong, C. Liu, X. Xie, E. Chen, and H. Xiong. Social marketing meets targeted customers: A typical user selection and coverage perspective. In *ICDM*, page 350–359, 2014.

[51] Y. Liu, Z. Ren, W.-N. Zhang, W. Che, T. Liu, and D. Yin. Keywords generation improves e-commerce session-based recommendation. In *WWW*, page 1604–1614, 2020.

[52] M. Mayrhofer, J. Matthes, S. Einwiller, and B. Naderer. User generated content presenting brands on social media increases young adults' purchase intention. *International Journal of Advertising*, 39(1):166–186, 2020.

[53] D. Mimno, H. Wallach, E. Talley, M. Leenders, and A. McCallum. Optimizing semantic coherence in topic models. In *EMNLP*, page 262–272, 2011.

[54] P. Mukherjee and B. J. Jansen. Performance analysis of keyword advertising campaign using gender-brand effect of search queries. *Electronic Commerce Research and Applications*, 13(2):139–149, 2014.

[55] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin. Automatic evaluation of topic coherence. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, page 100–108, 2010.

[56] H. T. Nguyen, T. N. Dinh, and M. T. Thai±. Cost-aware targeted viral marketing in billion-scale networks. In *INFOCOM*, pages 1–9, 2016.

[57] M. Röder, A. Both, and A. Hinneburg. Exploring the space of topic coherence measures. In *WSDM*, page 399–408, 2015.

[58] R. L. T. Santos, C. Macdonald, and I. Ounis. Search result diversification. *Foundations and Trends® in Information Retrieval*, 9(1):1–90, 2015.

[59] M. Saraswat and S. Chakraverty. Enriching topic coherence on reviews for cross-domain recommendation. *The Computer Journal*, 65(1):80–90, 2022.

[60] L. Seeman and Y. Singer. Adaptive seeding in social networks. In *FOCS*, pages 459–468, 2013.

[61] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.

[62] L. Sun, W. Huang, P. S. Yu, and W. Chen. Multi-round influence maximization. In *KDD*, pages 2249–2258, 2018.

[63] F. Tang, Q. Liu, H. Zhu, E. Chen, and F. Zhu. Diversified social influence maximization. In *ASONAM*, page 455–459, 2014.

[64] Y. Tang, X. Xiao, and Y. Shi. Influence maximization: Near-optimal time complexity meets practical efficiency. In *SIGMOD*, page 75–86, 2014.

[65] K. Theocharidis, P. Karras, M. Terrovitis, S. Skiadopoulos, and H. W. Lauw. Adaptive content-aware influence maximization via online learning to rank. *TKDD*, 18(6), 2024.

[66] K. Theocharidis and H. W. Lauw. Community similarity based on user profile joins. In *EDBT*, page 572–583, 2024.

[67] K. Theocharidis, M. Terrovitis, S. Skiadopoulos, and P. Karras. A content recommendation policy for gaining subscribers. In *SIGIR*, page 2501–2506, 2022.

[68] N. Ukey, Z. Yang, B. Li, G. Zhang, Y. Hu, and W. Zhang. Survey on exact knn queries over high-dimensional data space. *Sensors*, 23(2), 2023.

[69] N. Ukey, G. Zhang, Z. Yang, B. Li, W. Li, and W. Zhang. Efficient continuous knn join over dynamic high-dimensional data. *World Wide Web*, 26(6):3759–3794, 2023.

[70] S. Vaswani, B. Kveton, Z. Wen, M. Ghavamzadeh, L. V. S. Lakshmanan, and M. Schmidt. Model-independent online learning for influence maximization. In *ICML*, pages 3530–3539, 2017.

[71] C. Wartena, W. Slakhorst, and M. Wibbels. Selecting keywords for content based recommendation. In *CIKM*, page 1533–1536, 2010.

[72] Z. Wen, B. Kveton, M. Valko, and S. Vaswani. Online influence maximization under independent cascade model with semi-bandit feedback. In *NeurIPS*, pages 1–24, 2017.

[73] Q. Wu, Z. Li, H. Wang, W. Chen, and H. Wang. Factorization bandits for online influence maximization. In *KDD*, pages 636–646, 2019.

[74] C. Yang, X. Yu, and Y. Liu. Continuous knn join processing for real-time recommendation. In *ICDM*, page 640–649, 2014.

[75] D.-N. Yang, H.-J. Hung, W.-C. Lee, and W. Chen. Maximizing acceptance probability for active friending in online social networks. In *KDD*, pages 713–721, 2013.

[76] Y. Yao and F. M. Harper. Judging similarity: a user-centric study of related item recommendations. In *RecSys*, pages 288–296, 2018.

[77] W.-t. Yih, J. Goodman, and V. R. Carvalho. Finding advertising keywords on web pages. In *WWW*, pages 213–222, 2006.

[78] J. Yuan and S. Tang. No time to observe: Adaptive influence maximization with partial feedback. In *IJCAI*, page 3908–3914, 2017.

[79] Y. Zhang, J. Guo, W. Yang, and W. Wu. Targeted activation probability maximization problem in online social networks. *IEEE Transactions on Network Science and Engineering*, 8(1):294–304, 2020.

[80] Y. Zhu, Y. Lan, J. Guo, X. Cheng, and S. Niu. Learning for search result diversification. In *SIGIR*, page 293–302, 2014.