Content Sharing Design for Social Welfare in Networked Disclosure Game

Feiran Jia[1] **Chenxi Qiu**[2] **Sarah Rajtmajer**[1] **Anna Squicciarini**[1]

[1]Information Sciences and Technology, Pennsylvania State University,Pennsylvania, USA
[2]Computer Science and Engineering, University of North Texas,Texas, USA

Abstract

This work models the costs and benefits of personal information sharing, or self-disclosure, in online social networks as a networked disclosure game. In a networked population where edges represent visibility amongst users, we assume a leader can influence network structure through content promotion, and we seek to optimize social welfare through network design. Our approach considers user interaction non-homogeneously, where pairwise engagement amongst users can involve or not involve sharing personal information. We prove that this problem is NP-hard. As a solution, we develop a Mixed-integer Linear Programming algorithm, which can achieve an exact solution, and also develop a time-efficient heuristic algorithm that can be used at scale. We conduct numerical experiments to demonstrate the properties of the algorithms and map theoretical results to a dataset of posts and comments in 2020 and 2021 in a COVID-related Subreddit community where privacy risks and sharing tradeoffs were particularly pronounced.

1 INTRODUCTION

Online social engagement allows users to connect with peers and other contributors through discussion Brake [2012], Zafarani and Liu [2013], Golbeck [2007]. Frequently, online discussions entail *self-disclosure*, the voluntary sharing of personal information with others, which can include identifying or sensitive details such as location, age, gender, race, political affiliation, religious beliefs, and cognitive or emotional vulnerabilities Choi and Bazarova [2015]. Despite apparent privacy risks, acts of self-disclosure can enhance social rewards Hallam and Zanella [2017] by building trust, promoting empathy, increasing legitimacy and likeability, and deriving social support De Choudhury and De [2014], Seiter and Brophy [2021]. These trade-offs have become particularly evident over the past three years, during the Covid-19 pandemic Nabity-Grover et al. [2020], Blose et al. [2020], Umar et al. [2021], Amosun et al. [2021].

This work leverages a game-theoretic model to formalize tradeoffs between privacy risks and social rewards in an online social network. We define utility at the individual user level, and in doing so, we enable a notion of global *social welfare*. We study social welfare optimization through strategic network design, operationalized through content promotion, as a game-theoretic problem.

Our use of game theory is motivated by the view of self-disclosure (SD) as inherently social and strategic behavior. We posit that, implicitly or explicitly, users experience a payoff for sharing behaviors, which is mediated by gains and (privacy) losses. In parallel, a social network provider benefits from engaged communities, where users' gratification and sense of community are maximized. In particular, we structure this problem as a Stackelberg game Li and Sethi [2017] in which a leader moves first, and followers move sequentially thereafter. The OSN platform is the game's leader; it can significantly influence the pairwise interactions defining the social network graph. OSN users are followers in the game; each decides whether and what to disclose about themselves, given their network of peers and perceived costs and benefits of sharing.

We formalize this problem as a bi-level programming problem and prove that this problem is NP-hard. As a solution, we linearize the problem to a mixed integer linear programming problem, which can obtain the optimal solution for a small network (e.g., 32 users). In addition, we propose a Greedy naive solution and a time-efficient heuristic algorithm for deployment at scale. The algorithm dynamically maintains a rank of users based on their estimated utilities and removes the nodes and associating edges with the most negligible contribution to social welfare.

Finally, we put the networked disclosure game (NDG) into

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

practice through simulations using the traces collected from a real dataset of online conversations in a SubReddit for COVID-positive users. We first evaluate our algorithmic solutions' performance and computational efficiency using synthetic networks. Subsequently, we expand our experiments on real-world data, identifying parameter values and network criteria conducive to maximizing social welfare.

2 RELATED WORK

Self-disclosure. Self-disclosure has been studied as an intentional and influenced behavior and has both intrinsic and extrinsic rewards Abramova et al. [2017], Huang [2016], Wang et al. [2016]. A growing body of work looks at SD both in online communications and particularly in the context of social networks. Work has shown that SD can support psychological well-being Wingate et al. [2020], Amosun et al. [2021] and plays a role in building and maintaining relationships, social connectedness, and emotional support Barak and Gluck-Ofri [2007], Aharony [2016], Shih et al. [2015], Wingate et al. [2020], Amosun et al. [2021]. Increasingly, studies have explored self-disclosure in OSNs where users interact with others also seeking these benefits Lee et al. [2013], Andalibi et al. [2016], Huang [2016], Pan et al. [2018], Trepte et al. [2018], Wingate et al. [2020], Abramova et al. [2017], Hallam and Zanella [2017]. While several contributions have studied the impacts of self-disclosure in social networks, this has been mostly qualitative work or has relied on small datasets for quantitative observations Huang [2016], Hallam and Zanella [2017]. To our knowledge, few studies have proposed computational measurement of the social benefits of personal information sharing. In Griffin et al. [2019] authors proposed a public goods framework to measure self-disclosure benefits as a shared resource, where voluntary disclosure is framed as contributing to the richness of dialogue from which all participants benefit. This framing neglects individual privacy risks in the conceptualization of public good and is insufficient to capture the heterogeneity of individual disclosure decisions as they are influenced by peers. Our work attempts to fill this important gap by addressing the group-level social benefits of self-disclosure.

Of note, recent work also has focused on automated models to scale self-disclosure labels in small manually-annotated data to larger samples Bak et al. [2012, 2014], Houghton and Joinson [2012], Umar et al. [2019]. These models employ highly curated dictionaries and extensive feature engineering, which limit the inference process and performance on unseen data. Other recent studies have used transfer-learning techniques on NLP models for labeling self-disclosure in text Pant et al. [2020]. We employ state-of-the-art BERT-based models to label our dataset for analysis in this work. Our emphasis is not on detecting utterances of self-disclosure in online conversations but rather on the impact these messages have on the welfare of their broader community.

Social Influence. Social networks are extensively used for information diffusion, with concrete examples such as viral marketing Domingos and Richardson [2001] and targeted advertisement Li et al. [2011]. These applications have garnered significant attention from the social network analysis community, specifically, the topic of influence maximization Kempe et al. [2003]. At a high level, influence maximization aims to select a small subset of *seeds* acting as the source of influence with which the influence is maximized under some information diffusion model, e.g., Independent Cascade Wang et al. [2012], Linear Threshold Goyal et al. [2011], where the influence is measured by the number of affected nodes in the network. In contrast to a diffusion setting, Irfan and Ortiz [2014] propose the Linear Influence Game (LIG) as a simultaneous move game. Our inner networked disclosure game features analogous utility functions to LIG, albeit with binary actions confined to $\{0, 1\}$, positive weights, and undirected network structure. This enables us to obtain a polynomial time solution for identifying a Pure Strategy Nash Equilibrium (PSNE). Compared with the influence works, our primary goal is to promote content and design network structure, which are more similar to recommendation Corò et al. [2019], Coró et al. [2021].

3 PROBLEM FORMULATION

We assume the conversational context of an OSN. Namely, users interact and share "with" one another at a given time. These interactions can be represented as a network graph. We assume that all users within that context receive a (measurable) reward for participation. In practice, the measurement of reward is likely to be platform-dependent. Here, we consider measures of social engagement, including likes, shares, the sentiment of replies, and similar. We assume that the act of self-disclosing comes at a privacy cost to the individual user. This cost can be instantiated in various ways, i.e., proportional to the size of neighbors and other agents, and can be set to any [0,1] value Liu and Terzi [2010]. In our experiments, as we return in Section 5.2, this cost is assumed constant and learned from users' interactions with peers.

From these assumptions, we cast the problem as a binary networked game. Specifically, users who share personal information incur an individual cost to privacy but receive social or emotional rewards.

Central to this formulation is the notion of social welfare Zheleva and Getoor [2009], defined as the total utility of all participating users in the game, where utility is measured as the difference between cost and reward. From this perspective, we explore interventions, e.g., content ranking and recommendation, that OSN platforms might incorporate to

support greater social welfare.

Following, we introduce the mathematical models of (1) the *Networked Disclosure Game* and (2) the *Content Promotion Networked Design*.

Networked Disclosure Game (NDG). The *networked disclosure game (NDG)* is defined on a content promotion network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, ..., n\}$ are the users, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a reflection of whether two users can see each other's content (i.e. the interdependencies among the players' utilities). We use a binary decision variable $x_i \in \{0, 1\}$ $(i = 1, ..., N)$ to indicate user i's strategy to self-disclose, i.e., $x_i = 1$ if user i posts content containing self-disclosure; otherwise $x_i = 0$. We let $\mathbf{x} = (x_1, ..., x_N)$ be the whole action profile.

As noted above, users engage in self-disclosure to obtain benefits, e.g., emotional or informational support, broadly represented as social connectivity. Our model invokes the concepts proposed by social penetration theory, suggesting that mutual information sharing is a cornerstone for building relationships. In essence, users can reap the benefits from their neighbors only when they indulge in self-disclosure. Specifically, given a pair of users i and j sharing information with each other ($x_i = x_j = 1$, and $(i, j) \in \mathcal{E}$), the benefit that user i receives from user j is denoted by a weight $w_{i,j}$ from i to j[1]. As such, we calculate the total benefit that user i receives from his/her neighbors $\mathcal{N}_i = \{j|(j, i) \in \mathcal{E}\}$ as $g_i(x_i, \mathbf{x}_{-i}) = x_i \sum_{j \in \mathcal{N}_i} w_{j,i} x_j$, where $w_{j,i}$ denotes the potential impact user j has to user i, and $\mathbf{x}_{-i}$ denotes the strategy profiles except for user i.

On the other hand, self-disclosure comes at a potential privacy cost for a user, i.e., personal information that is voluntarily shared may result in increased privacy risks[2]. Therefore, we define the utility for each user i as

$$U_i(x_i, \mathbf{x}_{-i}|\mathcal{G}) = g_i(x_i, \mathbf{x}_{-i}|\mathcal{G}) - c_i x_i \tag{1}$$

where $c_i > 0$ is the cost of user i for disclosing (e.g. perceived privacy loss). To normalize the weights and the costs, we assume $\sum_j w_{j,i} \leq 1$ and $c_i \in [0, 1]$.

In this paper, we adopt the *pure strategy Nash Equilibrium (PSNE)* as our solution concept, i.e., $U_i(x_i, \mathbf{x}_{-i}|\mathcal{G}) > U_i(1 - x_i, \mathbf{x}_{-i}|\mathcal{G})\ \forall i$. Moreover, we assume that each user i breaks ties in favor of disclosing, i.e., user i chooses to disclose when

$$U_i(1, \mathbf{x}_{-i}|\mathcal{G}) \geq U_i(0, \mathbf{x}_{-i}|\mathcal{G}), \text{ or simply } \textstyle\sum_{j \in \mathcal{N}_i} w_{j,i} x_j \geq c_i. \tag{2}$$

where Equ. (2) is called the **threshold condition**.

Theorem 3.1. *The strategy profile* $\mathbf{x}$ *is a PSNE if and only if (1) everyone who invests satisfies the* ***threshold condition*** *and (2) other agents do not satisfy* ***threshold condition****, i.e.* $\sum_{j \in \mathcal{N}_i} w_{j,i} x_j < c_i$ *(Detailed proof can be found in the supplementary material).*

We define social welfare as the sum utility of all the users.

$$SW(\mathbf{x}|\mathcal{G}) = \textstyle\sum_{i \in \mathcal{V}} U_i(x_i, \mathbf{x}_{-i}|\mathcal{G}) \tag{3}$$

This formulation follows that of Yu et al. [2020], which similarly seeks to find social welfare-optimal equilibrium in the context of a binary networked public goods game.

Content Sharing Network Design. OSN platform providers control whose content to show (or promote) among a set of users $\mathcal{V} = \{1, ..., N\}$. The decision to connect or promote connections between two users i and j can be considered a recommendation. We let $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denote the set of recommended connections and $\lambda_{i,j}$ denote the cost of recommending each $e_{i,j} \in \mathcal{E}$.

Additionally, we assume the platform has a finite set of resources or a budget B to promote users' interactions, i.e., $\sum_{e_{i,j} \in \mathcal{E}} \lambda_{i,j} \leq B$, where the recommended connections are limited to an *input edge set* $\mathcal{E}^{\text{in}}$, i.e., $\mathcal{E} \subseteq \mathcal{E}^{\text{in}}$. For instance, in practice, the OSN's action space is restricted by its infrastructure, which might only have the capability to promote a limited set of users and their content. Our goal is to maximize social welfare by solving the *Optimal Self-Disclosure Sharing Problem (OSDSP)*, defined as follows.

Definition 3.2. *(OSDSP) Given an input graph* $\mathcal{G}^{in} = (\mathcal{V}, \mathcal{E}^{in})$*, the user's utility functions* $U_i(\cdot|\cdot)$*,* $\forall i$*, and a budget* $B \geq 0$*, OSDSP in one shot can be defined as finding an optimal edge set* $\mathcal{E} \subseteq \mathcal{E}^{in}$*,* $\sum_{e_{i,j} \in \mathcal{E}} \lambda_{i,j} \leq B$*, such that the PSNE of the NDG defined on* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *maximizes the social welfare. Here, when there are multiple PSNEs, only the one with the highest maximized social welfare is considered.*

Theorem 3.3. *The OSDSP problem is NP-hard.*

Proof. We prove this theorem by constructing a polynomial time reduction from the NP-complete problem *subset sum* problem Bhasin [2015] to OSDSP. Before the proof, we introduce the decision problems of both subset sum and OSDSP:

Definition: The decision problem of subset sum.
Instance: Given a set of M positive integers $\mathcal{A} = \{a_1, ..., a_M\}$ and a target sum value A.

[1]For simplicity, we disregard the impact of negative social interactions and assume that interactions generally provide some benefit. Note that since trust and influence are not always symmetric, $w_{j,i}$ and $w_{i,j}$ can be different values, connected with the benefit functions g_i and g_j separately

[2]Note that for the purposes of this model, risks may not be immediately tangible or easy to quantify. This is irrelevant since we model users' own decision process and perceived rather than objective risks.

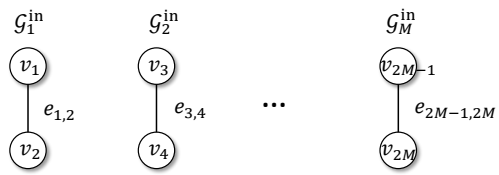

Figure 1: The OSDSP instance in the NP hard proof

Question: Whether exists a subset $\mathcal{A}' \subseteq \mathcal{A}$, such that the sum of the elements in $\mathcal{A}'$ is equal to A, i.e., $\sum_{a_i \in \mathcal{A}'} a_i = A$.

Definition: The decision problem of OSDSP.
Instance: Given an input network $\mathcal{G}^{\text{in}} = (\mathcal{V}, \mathcal{E}^{\text{in}})$, impact coefficients $\{w_{i,j}\}_{e_{i,j} \in \mathcal{E}^{\text{in}}}$, recommendation costs (edge costs) $\{\lambda_{i,j}\}_{e_{i,j} \in \mathcal{E}^{\text{in}}}$, cost coefficients $\{c_i\}_{i \in \mathcal{V}}$, a constant U, and a budget limit B.
Question: Whether exists a solution $(\mathbf{x}, \mathcal{G})$ such that $SW(\mathbf{x}|\mathcal{G}) \geq U$ and $\sum_{e_{i,j} \in \mathcal{E}} \lambda_{i,j} \leq B$.

Given any subset sum problem instance, we can construct the following OSDSP instance:
1) There are $2M$ users (nodes) in the network;
2) As Fig. 1 shows, $\mathcal{G}^{\text{in}}$ is composed of M disjoint sub-graphs $\mathcal{G}_1^{\text{in}}, ..., \mathcal{G}_M^{\text{in}}$, where each $\mathcal{G}_i^{\text{in}}$ is composed of two nodes v_{2i-1} and v_{2i} connecting by an edge $e_{2i-1,2i}$ with edge cost $\lambda_{2i-1,2i} = a_i$;
3) In each $\mathcal{G}_i^{\text{in}}$, the two nodes v_{2i-1} and v_{2i} have their costs $c_{2i-1} = c_{2i} = 1.5a_i$, and $w_{2i-1,2i} = w_{2i,2i-1} = 2a_i$;
4) $U = A$ and $B = A$.

This reduction process from subset sum to OSDSP is performed in polynomial time. Before showing the correctness of this reduction, we first give Lemma 3.1 as a preparation (the detailed proof can be found in the supplementary material):

Lemma 3.1. *For each sub-graph $\mathcal{G}_i^{in}$, there are only two possible PSNEs: (1) both nodes v_{2i-1} and v_{2i} self-disclose, or (2) neither nodes disclose itself.*

We show the correctness of the polynomial reduction, i.e., *a solution exists for the subset sum instance if only if there exists a feasible solution for the OSDSP instance*. Note that in Case (1) of Lemma 3.1, the total social welfare of $\mathcal{G}_i^{\text{in}}$ is $\frac{a_i}{2} + \frac{a_i}{2} = a_i$ and the total cost of the promoted edges is a_i. In Case (2), the total social welfare of $\mathcal{G}_i^{\text{in}}$ is 0 (no node disclosed itself), and the total promoted edge cost is 0.

$\Rightarrow$: Assuming exists a solution $\mathcal{A}' = \{a_{i_1}, a_{i_2}, ..., a_{i_m}\}$ for the subset sum instance, i.e., $\sum_{l=1}^{m} a_{i_l} = A$, we can construct a feasible solution of the OSDSP instance: For each sub-graph $\mathcal{G}_{i_l}^{\text{in}}$ ($l = 1, ..., m$), we promote the edge $e_{2i-1,2i}$ and let the nodes disclose, i.e., $x_{2i-1} = x_{2i} = 1$. The total social welfare in $\mathcal{G}_{i_l}^{\text{in}}$ is equal to $\sum_{l=1}^{m} a_{i_l} = A \geq U$ and the total cost $\sum_{l=1}^{m} a_{i_l} = A \leq B$.

$\Leftarrow$: Assuming exists a solution $(\mathbf{x}, \mathcal{G})$ in the OSDSP instance, where the edges (nodes resp.) in the sub-graphs $\mathcal{G}_{i_1}^{\text{in}}, \mathcal{G}_{i_2}^{\text{in}}, ..., \mathcal{G}_{i_m}^{\text{in}}$ are promoted (disclosed resp.). Hence,

$$SW(\mathbf{x}|\mathcal{G}) = \sum_{l=1}^{m} a_{i_l} \geq U = A \tag{4}$$

$$\sum_{e_{i,j} \in \mathcal{E}} \lambda_{i,j} = \sum_{l=1}^{m} a_{i_l} \leq B = A \tag{5}$$

indicating that $\sum_{l=1}^{m} a_{i_l} = A$. Hence $\mathcal{A}' = \{a_{i_1}, ..., a_{i_m}\}$ is a feasible solution of the subset sum instance. □

4 ALGORITHM DESIGN

Due to the hardness of OSDSP (Theorem 3.3), in this section, we aim to design algorithms that can achieve near-to-optimal solutions with high time efficiency.

OSDSP can be decomposed into two subproblems:
Inner problem - Given a promotion network $\mathcal{G}$, how to find an optimal PSNE $\mathbf{x}$ of NDG to maximize the social welfare. The inner problem establishes a relationship between the promotion network $\mathcal{G}$ and its maximum social welfare, denoted by $\sigma(\mathcal{G})$ (formally defined in Definition 4.3).
Outer problem - Given $\sigma(\mathcal{G})$, how to find the optimal promotion network $\mathcal{G}^*$ to maximize the social welfare while satisfying the budget limit. Like Coró et al. [2021], we assume that promoting each piece of content has an equal cost for simplicity since in reality the amount of online space each content occupies is similar when displayed. Then the outer problem can be formulated as $\mathcal{G}^* = \arg\max_{|\mathcal{E}| \leq B} \sigma(\mathcal{G})$.

To solve the inner problem, we design an iterative algorithm in Section 4.1 and prove that the algorithm can achieve the optimal PSNE. For the outer problem, we prove that it is super modular under certain conditions ($\frac{c_i}{w_{j,i}} \leq 1, \forall (i,j) \in \mathcal{E}^{in}$) in Section 4.2. Based on the insights obtained in Section 4.1 and Section 4.2, we design two time-efficient heuristic algorithms to solve OSDSP in Section 4.3. Finally, to evaluate how close the heuristic can achieve the optimal of OSDSP, in Section 4.4, we design a *mixed integer linear programming (MILP)* based method that can achieve the optimal solution in a relatively small scale, as a benchmark of the heuristic in Section 5.

4.1 INNER PROBLEM: SOLVING THE NDG

Given a promotion network, the NDG is likely to have multiple PSNEs. For example, $\mathbf{x} = \mathbf{0}$ is a PSNE, denoted as a trivial PSNE. We are interested in the optimal PSNE that maximizes social welfare in the inner problem.

To find the optimal PSNE, we design an iterative algorithm called the *MaxInvest* algorithm. The pseudo-code is shown

in Algorithm 1. Here, we use the superscript $^{(k)}$ to denote the values set/derived in iteration k.

The algorithm starts by initializing a *potential benefit value* $h_i^{(0)} = \sum_{j \in \mathcal{N}_i} w_{j,i} x_j^{(0)}$ for each node i ($i = 1, ..., n$) by assuming each node i discloses, i.e., $x_i^{(0)} = 1$ (line 2). Then the algorithm traverses each code i and deactivate it (by setting $x_i^{(0)} = 0$) when its threshold condition $h_i^{(0)} \geq c_i$ is not satisfied (line 3). Since deactivating nodes will also decrease their neighbors' potential benefit values, in the next part (line 4-12), we iteratively check the threshold conditions of the deactivated nodes' neighbors: We maintain a queue Q and in each iteration k, we pop the front node off the queue and deactivate it (line 6). We then check its neighbors' threshold conditions (line 10) and push the ones violating the conditions onto Q (line 11). This process is repeated until the queue is empty.

Algorithm 1 MaxInvest

1: **procedure** MAXINVEST($\mathcal{E}$)
2: **Initialization:** an empty queue Q, a potential benefit array $\left\{h_1^{(0)}, ..., h_n^{(0)}\right\}$, where $h_i^{(0)} = \sum_{j \in \mathcal{N}_i} w_{j,i} x_j$; a strategy profile $\mathbf{x}^{(0)} = \left\{x_1^{(0)}, ..., x_n^{(n)}\right\}$, where $x_i^{(0)} = 1, \forall i \in V$.
3: Check the *threshold condition*, i.e. $h_i^{(0)} \geq c_i$ for each node and push those who violate the condition onto Q;
4: Iteration index $k \leftarrow 1$;
5: **while** Q is not empty **do**
6: Pop the front node i off Q;
7: Deactivate node i by set $x_i^{(k)} = 0$;
8: **for** each node i's neighbour node j **do**
9: Update the potential benefit $h_j^{(k)} = h_j^{(k-1)} - w_{i,j}$;
10: **if** $h_j^{(k)} < c_j$ and $x_j^{(k-1)} = 1$ **then**
11: Push node j onto Q;
12: $k \leftarrow k + 1$;
13: **Return:** $\mathbf{x}^{(k)}$

Next, we prove that the users' action profile $\mathbf{x}$ returned by MaxInvest is both *feasible* (in Theorem 4.1), i.e., $\mathbf{x}$ is a PNSE, and *optimal* (in Theorem 4.2), i.e., it achieves the maximum social welfare.

Lemma 4.1. *Let $h_i^{(k)}$ denote h_i in the kth iteration in MaxInvest. For any iteration $k_1 < k_2$, we have $h_i^{(k_1)} \geq h_i^{(k_2)}$, i.e., the potential benefit of each node i is non-increasing over iterations in MaxInvest.*

Proof. $h_i^{(k_1)} - h_i^{(k_2)} = \sum_{j \in \mathcal{N}_i} w_{j,i} \left(x_j^{(k_1)} - x_j^{(k_2)}\right) \geq 0.$ □

Theorem 4.1. $\mathbf{x}$ *returned by MaxInvest is a PSNE.*

Proof. Let K be the total number of iterations of MaxInvest. We need to prove that (1) for any $x_i = 1$, $h_i^{(K)} \geq c_i$, and (2) for any $x_i = 0$, $h_i^{(K)} < c_i$.

(1) For the sake of contradiction, we assume that there exists $x_i = 1$ such that $h_i^{(K)} < c_i$. First, $h_i^{(0)} \geq c_i$; otherwise node i is removed at initialization. According to Lemma 4.1, there must exists an iteration k ($k \leq K - 1$) such that $h_i^{(k)} \geq c_i$ and $h_i^{(k+1)} < c_i$. The decrease of h_i at the iterations $k + 1$ has to be caused by the leave of at least one of i's neighbors, say node v_j. Then, when removing v_j, as v_j's neighbor, i's updated threshold condition has to be checked, which is $h_i^{(k+1)} < c_i$, and hence v_j should have been added to Q and removed by the algorithm terminates, which is contradicted with the assumption that $x_i = 1$.

(2) For each $x_i = 0$, we let k_i denote the iteration when node i is removed, which indicates that $h_i^{(k_i)} < c_i$. According to Lemma 4.1, $h_i^{(K)} \leq h_i^{(k_i)} < c_i$. The proof is completed. □

Lemma 4.2. *Suppose $\mathbf{x}$ is the profile returned by the MaxInvest. For any PSNE profile $\mathbf{x}'$, we have $\mathbf{x}' \leq \mathbf{x}$.*

Theorem 4.2. *(Optimality) Given $\mathcal{G}$, the strategy profile $\mathbf{x}$ returned by Algorithm 1 is a PSNE that maximizes the social welfare.*

4.2 ANALYSIS OF THE OUTER PROBLEM

In this part, we analyze the properties of the outer problem under certain conditions. Note that the MaxInvest algorithm for the inner problem has established the relationship between any promotion network $\mathcal{G}$ and the maximum social welfare it can achieve. Such a relationship can be described by the optimal social welfare function (Definition 4.3):

Definition 4.3. *We define the optimal social welfare function σ as a map from a given edge set $\mathcal{E} \subseteq \mathcal{E}^{in}$ to the social welfare of the profile $\mathbf{x}$ returned by the MaxInvest $\sigma(\mathcal{E}) = SW(\mathbf{x}|(\mathcal{V}, \mathcal{E}))$, where $\mathbf{x} = MaxInvest(\mathcal{E})$.*

Accordingly, the outer problem of OSDSP can be rewritten as:

$$\max\ \sigma(\mathcal{E}) \quad \text{s.t. } |\mathcal{E}| \leq B. \tag{6}$$

Note that we don't have the closed form of the function $\sigma(\cdot)$, which can be only evaluated by running the MaxInvest algorithm (Algorithm 1). Next, we first prove that $\sigma(\mathcal{E})$ is *monotonic* (Theorem 4.5) and super-modular when $\frac{c_i}{w_{j,i}} \leq 1, \forall (i,j) \in \mathcal{E}^{in}$ (Theorem 4.6).

Definition 4.4. *We define the optimal investment function of a given edge set $\mathcal{E} \subseteq \mathcal{E}^{in}$ as the number of the invest agents of the profile $\mathbf{x}$ returned by the MaxInvest $I(\mathcal{E}) = \sum_{i \in \mathcal{V}} x_i$, where $\mathbf{x} = MaxInvest(\mathcal{E})$.*

Theorem 4.5. *(Monotonicity) For all pairs of the edge sets $\mathcal{S}$ and $\mathcal{T}$ such that $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{E}^{(in)}$, we have (1) $MaxInvest(\mathcal{S}) \leq MaxInvest(\mathcal{T})$, (2) $I(\mathcal{S}) \leq I(\mathcal{T})$,and (3) $\sigma(\mathcal{S}) \leq \sigma(\mathcal{T})$.*

Theorem 4.6. *The optimal social welfare function $\sigma(\mathcal{E})$ is super-modular when $\frac{c_i}{w_{j,i}} \leq 1, \forall (i,j) \in \mathcal{E}^{in}$.*

Proof. We prove that for any edge $e^* \in \mathcal{E}^{(in)}$, and all pairs of the set $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{E}^{(in)}$, $\sigma(\cdot)$ satisfies $\sigma(\mathcal{S} \cup \{e^*\}) - \sigma(\mathcal{S}) \leq \sigma(\mathcal{T} \cup \{e^*\}) - \sigma(\mathcal{T})$. Detailed proof and general cases for both the optimal investment function and social welfare function are presented in the supplementary material. □

4.3 HEURISTIC ALGORITHMS OF THE OUTER PROBLEM

Algorithm 2 Greedy($\mathcal{G}^{in} = (\mathcal{V}, \mathcal{E}^{in}), B$)

1: Initialize $\mathcal{E} = \emptyset$;
2: **while** $|\mathcal{E}| \leq B$ **do**
3: **for all** edge $e \in \{\mathcal{E}^{in} \backslash \mathcal{E}\}$ **do**
4: $\Delta_e = \sigma(\mathcal{E} \cup \{e\}) - \sigma(\mathcal{E})$;
5: $\mathcal{E} = \mathcal{E} \cup \{e_* = argmax_e \Delta_e\}$;
6: **Output** $\mathcal{E}$

Based on the insights obtained from Section 4.1 and Section 4.2, in this section, we provide two heuristic algorithms to solve OSDSP.

Greedy. The first naive heuristic is the Greedy algorithm, of which the pseudo-code is shown in Algorithm 2. The algorithm initializes the promoted edge set $\mathcal{E}$ by empty (line 1), and then greedily selects the edge with the highest SW marginal gain and adds it to $\mathcal{E}$ (line 2–5) until the number of promoted edges reaches the budget.

RankHeuristic (Rank). The pseudo-code of the second heuristic, RankHeuristic, is shown in Algorithm 3. The basic idea of the algorithm is to first initialize the promoted edge set $\mathcal{E}$ by $\mathcal{E}^{in}$ (line 2), use MaxInvest to obtain the optimal action profile $\mathbf{x}$ (line 3), and rank the active nodes based their U_i (line 4). After that, the algorithm iteratively deactivates the nodes with the least potential contribution to social welfare (lines 6-7) and the edges directed from the nodes (lines 8-9) until the budget is satisfied, i.e., $|\mathcal{E}| \leq B$.

4.4 A MIXED INTEGER LINEAR PROGRAMMING

We offer an optimal solution (at a relatively small scale) using the MILP framework. We let the indicator variable $p_{i,j}$ denote whether the connection between users i and j is

Algorithm 3 RankHeuristic

1: **procedure** RANK($\mathcal{G} = (\mathcal{V}, \mathcal{E}^{\text{in}}), B$)
2: **Initialization:** $\mathcal{E} = \mathcal{E}^{\text{in}}$;
3: $\mathbf{x} = \text{MaxInvest}(\mathcal{E})$;
4: Sort the active nodes (with $x_i = 1$, $U_i > 0$) in ascending order according to U_i (the ordered users are denoted by as Q);
5: Remove the edges connecting the removed nodes from $\mathcal{E}$;
6: **while** $|\mathcal{E}| > B$ and Q is not empty **do**
7: Remove the node i with least utility in Q;
8: **for** i's neighbor j **do**
9: Remove $e_{i,j}$ from $\mathcal{E}$;
10: $\mathbf{x} = \text{MaxInvest}(\mathcal{E})$;
11: Compute the utility of the remaining users;
12: Order the nodes based on their utilities update Q;
13: **Return:** $\mathcal{E}$

promoted, i.e., if $e_{i,j} \in \mathcal{E}$, $p_{i,j} = 1$; otherwise $p_{i,j} = 0$. As we assume visibility to be symmetric, i.e. $\mathcal{E}$ is undirected), we have the constraints $p_{i,j} = p_{j,i}$, $\forall i, j \in \mathcal{V}$. We let $\mathbf{P} = \{p_{i,j}\}_{n \times n}$. The mixed integer programming (MIP) version of OSDSP can be formulated as

$$\max_{\mathcal{P}} \quad SW(\mathbf{x}, \mathbf{P}) = \sum_{i \in \mathcal{V}} x_i \sum_{e_{j,i} \in \mathcal{E}^{\text{in}}} w_{j,i} x_j p_{j,i} \tag{7}$$

$$\text{s.t.} \quad x_i \in \arg\max_{x_i \in \{0,1\}} x_i \sum_{e_{j,i} \in \mathcal{E}^{\text{in}}} w_{j,i} x_j p_{j,i}, \forall i \tag{8}$$

$$\sum_{e_{i,j} \in \mathcal{E}^{\text{in}}} \lambda_{i,j} p_{i,j} \leq B \tag{9}$$

$$x_i \in \{0,1\}, \ \forall i \in \mathcal{N}, p_{i,j} \in \{0,1\}, \ \forall e_{i,j} \in \mathcal{E}^{\text{in}} \tag{10}$$

The constraint Equ. (8) can be replaced by Equ. (11):

$$x_i \textstyle\sum_j x_j w_{j,i} p_{j,i} - c_i x_i \geq (1 - x_i) \sum_j x_j w_{j,i} p_{j,i} - c_i (1 - x_i) \tag{11}$$

i.e., user i can achieve a higher utility when choosing x_i than choosing $1 - x_i$, which reduces the bi-level structure of OSDSP to a single-level MIP problem. Here, we can linearize Equ. (11) by introducing intermediate variables $m_{i,j}$ for $\forall i, j \in \mathcal{V}$, such that:

$$0 \leq m_{j,i} \leq M x_j \text{ and } w_{j,i} p_{j,i} - M(1 - x_i) \leq m_{j,i} \leq w_{j,i} p_{j,i}. \tag{12}$$

Accordingly, each $p_{j,i} x_j$ in Equ. (11) can be replaced by $m_{j,i}$ as

$$m_{j,i} = \begin{cases} 0 & \text{if } x_j = 0 \\ w_{j,i} p_{j,i} & \text{if } x_j = 1 \end{cases} \Rightarrow m_{j,i} = w_{j,i} p_{j,i} x_j.$$

Then we linearize the non-linear term $x_i m_{i,j}$ by introducing variables $v_i = x_i (\sum_j m_{j,i}) \geq 0, \ \forall i \in \mathcal{V}$ and a big positive

number M. The following constraints should be satisfied :

$$\sum_j m_{j,i} - M(1 - x_i) \leq v_i \leq \sum_j m_{j,i}, \forall i \quad (13)$$

$$\text{and } 0 \leq v_i \leq M x_i, \forall i \quad (14)$$

Consequently, OSDSP can be formulated as a MILP:

$$\max \quad SW(\mathbf{x}|\mathcal{G}) = \sum_{i \in \mathcal{V}} SW_i(\mathbf{x}, \mathbf{P}) \quad (15)$$

$$\text{s.t.} \quad \text{Constraints}(9) - (14) \quad (16)$$

5 EXPERIMENTS

We conduct extensive experiments on both (1) synthetic networks and (2) real-world data representing users' online conversations during the Covid-19 crisis. For the latter set of experiments, we collected and labeled a rich dataset from the Reddit community, and following, we discuss our results in the context of that dataset. Our experiments are implemented in Python. All experiments have been performed on an Intel(R) Core(TM) i9-9820X CPU @ 3.30GHz. [3]

5.1 EXPERIMENTS ON SYNTHETIC DATA

This section presents our results for the following algorithmic solutions to solve the OSDSP problem. (1) **Betweeness:** As the simplest baseline, we consider an algorithm that selects edges based on the descending betweenness centrality. A node's betweenness is the number of shortest paths from every pair of nodes that pass through the node. (2) **Greedy:** The greedy algorithm described in Algorithm 2. (3) **MILP:** We use the optimization tool CPLEX IBM [2021] to compute the exact solution formulated in Eq. (15). (4) **Rank:** The RankHeuristic algorithm described in Algorithm 3.

We evaluate the performance of each algorithmic solution on the following three types of synthetic networks, $\mathcal{G} = (\mathcal{N}, \mathcal{E}^{in})$: (1) Complete networks, which implies the full rights of the network operator to promote contents (link) for pair of nodes; (2) Erdős-Rényi networks Erdös and Rényi [1959], which choose each of the possible edges with a given probability; (3) Barabasi-Albert (BA) networks Barabási and Albert [1999], scale-free networks with power-law degree distribution, and common in the online world. In our network generation process, 3 edges are attached to a new node from existing nodes. We present our findings for the optimality of the solutions and run-time.

Performance Analysis. We set the influence weight between agents as $w_{i,j} = w = 1/N, \forall i, j \in \mathcal{V}$. We denote the user's cost-weight ratio as $\gamma = \frac{c}{w}$. We assume that a node has probability η to have $\gamma \leq 1$ (private cost is set to $c \sim U(0, w)$) and probability $1 - \eta$ to be ($c \sim U(w, N)$). The budget B is set to $B = bN$, where b is denoted as a budget factor, and N is the number of nodes in the network.

We show social welfare results with varying η parameters in the 32-node BA and Complete network with 20 random seeds. In Figure 2 and Figure 3, the mean social welfare increases when η increases from 0 to 1. We can see that the Rank algorithm approximates the optimal solution well when η is low, whereas the Greedy algorithm shows better performance when facing the high η. This is because (1) when η is low, the Rank algorithm efficiently removes the edges connecting to nodes that are expensive and hard to incentivize; and (2) when η is high, the degree of supermodularity of the $\sigma(\cdot)$ is bounded, which benefits the Greedy algorithm.

Naturally, a higher promoting budget is beneficial to promote social welfare. Figure 4 demonstrates the overall social welfare with the budget factor b changed from 1.0 ($B = 32$) to 4.0 ($B = 128$). The optimality gap between the Greedy algorithm and MILP optimal solution is generally enlarged when the budget increases, whereas the Rank heuristic shows a stable and good performance.

Finally, we tune the edge creation probability ($0.3, 0.5, 0.7, 1.0$) of ER networks and discuss the effect of the search space (implies the size of $\mathcal{E}^{in}$). Figure 5 demonstrates that we can find better MILP and Greedy solutions resulting in higher social welfare when we have larger $\mathcal{E}^{in}$. However, the Rank heuristic may have a more significant optimality gap when the edge design space is enlarged.

Run time Analysis. To test the scalability of this method, we generate networks (Complete, BA, ER) of multiple node sizes and generate the ER networks with different generation probabilities.

Figure 6 shows the average running time in the log scale of the algorithms in Complete networks with $b = 3$ by tuning the size of network $N = 16, 20, 24, 32$. We can see that the average running time of MILP achieves more than 10^4 secs when it goes to 32 nodes.

We also examine the running time regarding η in Figure 7. When η increases, the average running time of MILP is significantly affected. Note that the run-time of Greedy, Betweenness, and Rank are stable as excepted (variances are tiny), so we do not plot the error bars. However, the running time of MILP has a huge variance. We can find some cases with incredibly long running time when $\eta = 1.0$, shown in Figure 8.

Although the MILP method provides an exact solution, it does not scale when the network has hundred of nodes (or more), and neither can greedy algorithms. One idea to reduce the time complexity is to cut the input edge space $\mathcal{E}^{in}$. Figure 9 shows the running time when $\mathcal{E}^{in}$ is created using a smaller probability, trading off the performance (Figure 5).

[3]The code is available at https://github.com/jtongxin/CSD-NDG

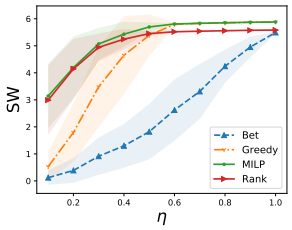

Figure 2: SW v.s. η of Complete networks, $B = 96$ over total $|\mathcal{E}^{in}| = 1024$

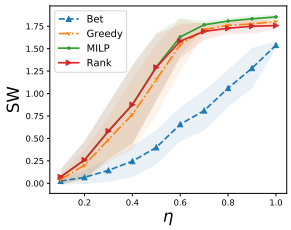

Figure 3: SW v.s. η of BA networks, $B = 32$ over total $|\mathcal{E}^{in}| = 87$

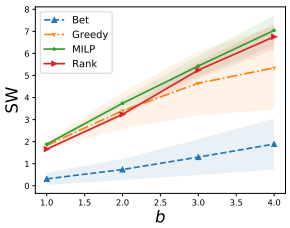

Figure 4: SW v.s. budget factor in Complete networks, $\eta = 0.5$

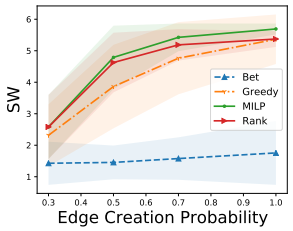

Figure 5: SW v.s. Edge Creation Probability in ER networks. $\eta = 0.5, b = 3.0$

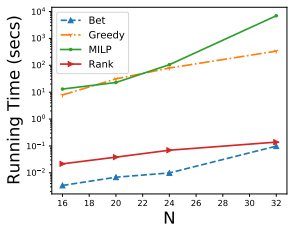

Figure 6: Running time (mean) v.s. N in the Complete networks.

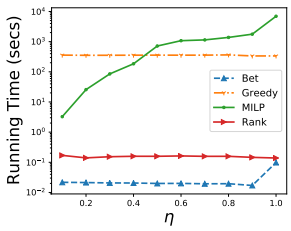

Figure 7: Running time (mean) v.s. η in the 32-node Complete networks.

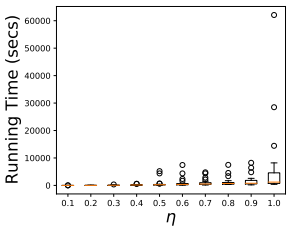

Figure 8: Running time of the MILP solutions v.s. η in Complete Networks.

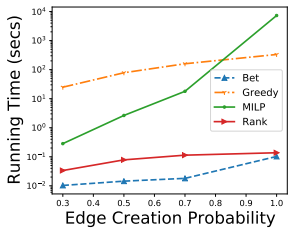

Figure 9: Running time (mean) v.s. Edge Creation Prob. in ER networks.

5.2 EXPERIMENTS ON REAL DATA

Data Set. Our dataset represents posts and comments collected from the Reddit online conversation platform. We collected data from the Subreddit community "CovidPositive" at three distinct moments during the pandemic. We consider 5629 posts and 50,526 associated comments from 15,172 unique users. For each comment or post, we collect the timestamp, message text, author id, and which reply the text is to (if any).

The three-part dataset represents selected posts and related comments collected from August 2020, September 2020, and April 2021. These periods were selected to capture snapshots of pandemic-centered conversations at different stages. In particular, April 2021 is seen as an important counterpoint to the August and September 2020 data, coinciding with the initial widespread availability of the COVID-19 vaccine. For each of the three periods, we create a representative social network graph (shown in Supplementary Material) where nodes are unique users and weighted, directed edges represent pairwise interactions between users in the form of a reply to a post or comment. Over all three months, the number of high-SD users totals more than half of all users in each network. This is an exceptional finding based on prior literature and one we suggest is connected with the particularly sensitive nature of the Subreddit. Descriptive statistics for each month are given in Table 1. Subreddit is most active in terms of user participation in August 2020.

Label Generation. We leverage BERT-based (pre-trained using bidirectional transformers Devlin et al. [2019]) approach to identify self-disclosure instances in our dataset. BERT's contextualized word representation and high labeling accuracy make it a suitable choice for this task, with superior performance compared to other existing techniques.

Our training dataset is the OffMyChest conversation dataset used for a self-disclosure detection task developed as part of the AFFCON 2020 Shared Task Jaidka et al. [2020]. The original dataset consists of 12,860 labeled sentences and 5,000 unlabeled sentences sampled from comments on subreddits within the OffMyChest community. We use the label of informational disclosure, emotional disclosure and do not distinguish between them in our analysis

We then utilized the uncased pre-trained BERT model to fine-tune for annotation, as proposed in Devlin et al. [2019]. Our loss function for this task was binary cross-entropy. Precisely, we use 80% of the Affcon training data for training and the remaining 20% for validation. We use an early stop strategy in training: we stop training when the validation loss does not decrease in 2 epochs. We choose the model that has the smallest loss on the validation set. Our F-1 score is 0.76, (precision 0.90, recall 0.66). We use a batch size of 16, max token size 200, for 10 epochs. Our learning rate is set as 2e-5. We obtained a 0.77 F1-score for information disclosure (precision 0.92, recall 0.66). The F1-score for emotional disclosure is 0.75 (precision 0.87, recall 0.65). This is consistent with the best-known performing model for emotional and information disclosure annotations on this dataset Pant et al. [2020]. We obtain 27155 self-disclosed

Month	Nodes	Edges	Weighted degree	Number of SCC	Size of largest SCC	Modularity
August 2020	6786	18688	3.3	3476	3139	0.528
September 2020	3305	7950	2.9	1657	1518	0.531
April 2021	4533	13013	3.5	2273	2164	0.605

Table 1: Network statistics for the three months of SubReddit data. SCC represents a strongly connected component. Modularity is calculated using the Louvain algorithm on the undirected network.

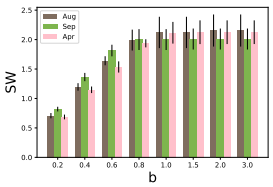

Figure 10: SW v.s. budget ratio in real data

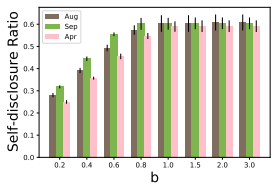

Figure 11: SD ratio v.s. budget ratio in real data.

(either emotional or informational) sentences out of a total of 50526.

Algorithmic Results. Given the reply network structures and Reddit dataset of August 2020, September 2020, and April 2021 as a starting point, we build the network $\mathcal{G}^{in} = (\mathcal{V}, \mathcal{E}^{in})$ as the undirected version of the reply network, and the user's utilities functions $U_i, \forall i \in \mathcal{V}$ as follows. (1) $w_{i,j}$: The influence weight coefficient of from agent i to j, proportion to the number of replies, denoted as $r_{j,i}$ from user j to user i in the reply network. We normalize $w_{i,j}$ by dividing the $\max_{v \in \mathcal{N}} \sum_u r_{v,u}$. In this way, $\sum_i w_{i,j} \leq 1$ is guaranteed. (2) c_i: The privacy cost coefficient of each user i. If the agent disclosed in the real data ($x_i^{\text{real}} = 1$), we sample the costs c_i from $U(0, h_i)$, where $h_i = \sum_j w_{j,i} x_j^{\text{real}}$. Otherwise, $c_i \sim U(h_i, 1)$. We calculate the coefficients c_i based on actual self-disclosure for both i's posts and comments. If the node has ever disclosed in any post or comments during the period, we consider $x_i^{\text{real}} = 1$.

We use the RankHeuristic algorithm to simulate the content promotion and NDG over the networks, varying the budget ratio $b = B/N$ from 0.2 to 3.0. Each experiment averages over 20 random seeds. Figure 10 compares the overall social welfare over the three months. We can observe that with the increase of the budget limit, the overall social welfare increases monotonically when $b < 1.0$ in all three months. When considering overall self-disclosure, many users are associated with low privacy cost coefficients. Thus, RankHeuristic can find an edge set that can induce saturated social welfare within a relatively small budget.

As reported in Figure 11, we also find that the self-disclosure ratio achieved by our algorithm in September is high, and the ratio in April is low when the budget is low. When we increase the budget, the gap between them is shrunk.

Due to the limited space, the computation time of RankHeuristic for the three months is listed in Table 1 in the supplementary file.

6 CONCLUSIONS AND DISCUSSIONS

In this work, we have developed a framework to rigorously model the impact of users' decisions in sharing personal information in online communities, wherein these individuals seek social reward, e.g., emotional support during a crisis, and in doing so, incur a cost to privacy. The presented theoretical results enable modeling social welfare on a directed network of interacting users. Critically, our approach allows us to find social network structures that optimize social welfare within platform constraints while respecting individual users' heterogeneous privacy preferences. Our research can guide the development of effective practices for ranking and recommending content on platforms, particularly in online communities that prioritize social support and connections despite potential privacy risks.

Despite the merits of this work, we acknowledge its limitations. A limitation of our work is precise mapping to the complexity of real-world scenarios and behaviors. Relatedly, it is knowingly difficult to quantify users' privacy preferences or individual benefits. Hence, fully validating this work is an important next step. Besides, the potential impact of fairness on content sharing, specifically the potential for biased treatment resulting in unequal access to opportunities for content promotion and shared information, has not been thoroughly examined when maximizing social welfare. Addressing fairness issues is a crucial aspect of future work.

Additional work should explore dependencies between social contexts where self-disclosure and support behaviors may be differently aligned, for instance, considering users' selfish behavior and negative/non-supportive interactions. Follow-on studies might also consider inducing the desired equilibrium Yu et al. [2021] to incentivize social welfare convergence at a reasonable pace. Finally, future work could fit this model to other datasets, varying in size, membership, audience, and topical focus.

Acknowledgements

This work was partially funded by the National Science Foundation under grant 2027757.

References

O. Abramova, A. Wagner, H. Krasnova, and P. Buxmann. Understanding self-disclosure on social networking sites-a literature review. Technical report, Darmstadt Technical University, Department of Business Administration, Economics and Law, Institute for Business Studies (BWL), 2017.

N. Aharony. Relationships among attachment theory, social capital perspective, personality characteristics, and facebook self-disclosure. *Aslib Journal of Information Management*, 2016.

T. S. Amosun, C. Jianxun, O. H. Rufai, S. Muhideen, R. Shahani, Z. Shah, and J. Koroma. Wechat usage during covid-19 pandemic lockdown: the mediating role of online self-disclosure on quality of friendship and well-being. *Global Knowledge, Memory and Communication*, 2021.

N. Andalibi, O. L. Haimson, M. De Choudhury, and A. Forte. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proc. of the 2016 CHI conference on human factors in computing systems*, pages 3906–3918, 2016.

J. Bak, C.-Y. Lin, and A. Oh. Self-disclosure topic model for classifying and analyzing twitter conversations. In *Proc. of EMNLP*, pages 1986–1996, Doha, Qatar, Oct. 2014. Association for Computational Linguistics.

J. Y. Bak, S. Kim, and A. Oh. Self-disclosure and relationship strength in twitter conversations. In *Proc. of ACL*, pages 60–64, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

A. Barak and O. Gluck-Ofri. Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior*, 10(3):407–417, 2007.

H. Bhasin. *Algorithms: Design and Analysis*. Oxford Univ Press, 2015.

T. Blose, P. Umar, A. Squicciarini, and S. Rajtmajer. Privacy in crisis: A study of self-disclosure during the coronavirus pandemic. *arXiv preprint arXiv:2004.09717*, 2020.

D. R. Brake. Who do they think they're talking to? framings of the audience by social media users. *International journal of communication*, 6:21, 2012.

Y. H. Choi and N. N. Bazarova. Self-disclosure characteristics and motivations in social media: Extending the functional model to multiple social network sites. *Human Communication Research*, 41(4):480–500, 2015.

F. Corò, G. D'Angelo, and Y. Velaj. Recommending links to maximize the influence in social networks. In *Proc. of IJCAI*, volume 4, pages 2195–2201. AAAI Press/IJCAI, 2019.

F. Coró, G. D'angelo, and Y. Velaj. Link recommendation for social influence maximization. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(6):1–23, 2021.

M. De Choudhury and S. De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*, 2014.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

P. Domingos and M. Richardson. Mining the network value of customers. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 57–66, 2001.

P. Erdös and A. Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290, 1959.

J. Golbeck. The dynamics of web-based social networks: Membership, relationships, and change. *First monday*, 2007.

A. Goyal, W. Lu, and L. V. Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *2011 IEEE 11th international conference on data mining*, pages 211–220. IEEE, 2011.

C. Griffin, S. Rajtmajer, P. Umar, and A. Squicciarini. Power law public goods game for personal information sharing in news commentaries. In *International Conference on Decision and Game Theory for Security*, pages 184–195. Springer, 2019.

C. Hallam and G. Zanella. Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards. *Computers in Human Behavior*, 68:217–227, 2017.

D. J. Houghton and A. N. Joinson. Linguistic markers of secrets and sensitive self-disclosure in twitter. In *2012 45th Hawaii International Conference on System Sciences*, pages 3480–3489, Jan 2012. doi: 10.1109/HICSS.2012.415.

H.-Y. Huang. Examining the beneficial effects of individual's self-disclosure on the social network site. *Computers in human behavior*, 57:122–132, 2016.

IBM. Decision optimization, October 2021. https://www.ibm.com/analytics/decision-optimization.

M. T. Irfan and L. E. Ortiz. On influence, stable behavior, and the most influential individuals in networks: A game-theoretic approach. *Artificial Intelligence*, 215:79–119, 2014.

K. Jaidka, I. Singh, J. Liu, N. Chhaya, and L. Ungar. A report of the cl-aff offmychest shared task: Modeling supportiveness and disclosure. In *AffCon@ AAAI*, pages 118–129, 2020.

D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

K.-T. Lee, M.-J. Noh, and D.-M. Koo. Lonely people are no longer lonely on social networking sites: The mediating role of self-disclosure and social support. *Cyberpsychology, Behavior, and Social Networking*, 16(6):413–418, 2013.

F.-H. Li, C.-T. Li, and M.-K. Shan. Labeled influence maximization in social networks for target marketing. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 560–563. IEEE, 2011.

T. Li and S. P. Sethi. A review of dynamic stackelberg game models. *Discrete & Continuous Dynamical Systems-B*, 22(1):125, 2017.

K. Liu and E. Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Trans. Knowl. Discov. Data*, 5(1):6:1–6:30, Dec. 2010. ISSN 1556-4681. doi: 10.1145/1870096.1870102. URL `http://doi.acm.org/10.1145/1870096.1870102`.

T. Nabity-Grover, C. M. Cheung, and J. B. Thatcher. Inside out and outside in: How the covid-19 pandemic affects self-disclosure on social media. *International Journal of Information Management*, 55:102188, 2020.

W. Pan, B. Feng, and V. Skye Wingate. What you say is what you get: How self-disclosure in support seeking affects language use in support provision in online support forums. *Journal of Language and Social Psychology*, 37 (1):3–27, 2018.

K. Pant, T. Dadu, and R. Mamidi. Bert-based ensembles for modeling disclosure and support in conversational social media text. In *AffCon@ AAAI*, pages 130–139, 2020.

C. R. Seiter and N. S. Brophy. Social support and aggressive communication on social network sites during the covid-19 pandemic. *Health communication*, pages 1–10, 2021.

Y.-W. Shih, M.-H. Hsu, and D.-C. Lee. Self-disclosure, interpersonal relationships, and stickiness of online communities. *Psychology and Behavioral Sciences*, 4(2):71, 2015.

S. Trepte, P. K. Masur, and M. Scharkow. Mutual friends' social support and self-disclosure in face-to-face and instant messenger communication. *The Journal of social psychology*, 158(4):430–445, 2018.

P. Umar, A. Squicciarini, and S. Rajtmajer. Detection and analysis of self-disclosure in online news commentaries. In *Proc. of WWW*, pages 3272–3278, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6674-8.

P. Umar, C. Akiti, A. Squicciarini, and S. Rajtmajer. Self-disclosure on twitter during the covid-19 pandemic: A network perspective. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 271–286. Springer, 2021.

C. Wang, W. Chen, and Y. Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25(3):545–576, 2012.

Y.-C. Wang, M. Burke, and R. Kraut. Modeling self-disclosure in social networking sites. In *Proc. of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 74–85. ACM, 2016.

V. S. Wingate, B. Feng, C. Kim, W. Pan, and J. Jang. The influence of self-disclosure in online support seeking on quality of received advice. *Journal of Language and Social Psychology*, 39(3):397–413, 2020.

S. Yu, K. Zhou, J. Brantingham, and Y. Vorobeychik. Computing equilibria in binary networked public goods games. In *Proc. of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2310–2317, 2020.

S. Yu, D. Kempe, and Y. Vorobeychik. Altruism design in networked public goods games. *arXiv preprint arXiv:2105.00505*, 2021.

R. Zafarani and H. Liu. Connecting users across social media sites: a behavioral-modeling approach. In *Proc. of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 41–49, 2013.

E. Zheleva and L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proc. of the 18th international conference on World wide web*, pages 531–540. ACM, 2009.