Incentive Design for Heterogeneous User-Generated Content Networks

[Extended Abstract]

Jie Xu
Electrical Engineering Department
University of California, Los Angeles
jiexu@ucla.edu

Mihaela van der Schaar
Electrical Engineering Department
University of California, Los Angeles
mihaela@ee.ucla.edu

ABSTRACT

This paper designs rating systems aimed at incentivizing users in UGC networks to produce content, thereby significantly improving the social welfare of such networks. We explicitly consider that monitoring user's production activities is imperfect. Such imperfect monitoring will lead to undesired rating drop of users, thereby reducing the social welfare of the network. The network topology constraint and users' heterogeneity further complicates the optimal rating system design problem since users' incentives are complexly coupled. This paper determines optimal recommendation strategies under a variety of monitoring scenarios. Our results suggest that, surprisingly, allowing a certain level of freeriding behavior may lead to higher social welfare than incentivizing all users to produce.

Categories and Subject Descriptors

J.4 [**Social and Behavior Sciences**]: Economics

General Terms

Theory

Keywords

User-Generated Content Networks, Incentives, Rating, Imperfect Monitoring

1. INTRODUCTION

Recent years have witnessed the rapid growth of user-generated content (UGC) networks where individuals establish connections with others to produce and share information, content and resources (e.g. Facebook, Twitter). Users obtain benefits from receiving the content produced by the users with which they are connected and incur costs when they produce content by themselves. (Fig. 1 provides an illustration of a UGC network.) Whether a UGC network can survive and thrive largely depends on the willingness of its users to produce and share content with other users within the network. However, it is observed in many studies (e.g. [1]) that users prefer freeriding, i.e. passively consuming content (e.g. by reading posts, watching videos shared by others) from others while producing no content, over actively producing content.

In this paper, we design rating systems for incentivizing users in a UGC network to produce and share content. The rating system is implemented and operated by the service or network administrator. Specifically, the administrator selects a social strategy to users which recommends to users levels of content production. For instance, a social strategy may recommend to all its users to produce content or it may recommend to only a subset of users to do so, while allowing freeriding from the rest of users. Depending on each user's compliance to the recommended social strategy, the rating of each user is updated over time (e.g. the rating of a complying user is increased and the rating of a non-compliant user is decreased). A key incentive for a user to maintain a high rating is that a corresponding content access level is imposed on users by the administrator depending on their ratings. For instance, low-rated users may have only limited access to the content produced by the users with which they are connected. This limited content access thus serves as a punishment device for the lurkers. We note that this kind of rating scheme can be easily implemented in most UGC networks. For example, on Facebook, a user can receive only a portion of their friends' wall posts.

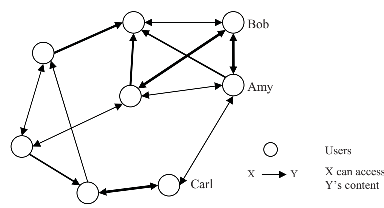

Figure 1: An illustrative UGC network.

We emphasize that monitoring is imperfect in practice and as a result, rating update errors are inevitable. For example, even though a user produced content, the system could mistakenly determine the user was freeriding and hence, its rating drops. This paper explicitly takes into consideration of the *imperfect monitoring* when designing the rating system. A distinguishing characteristic of UGC networks is that users are heterogeneous in terms of different connectivity and different benefits by receiving the same content. The problem becomes even more subtle since users are connected with others and hence, their incentives and resulting decisions are coupled with other's incentives and decisions. Therefore, one rating system design may provide sufficient incentives for some users to produce, but may not be sufficient for other users and one user choosing to lurk may influence the behavior of the users that connect to it since their benefits are also reduced due to this freeriding behavior. Unlike existing works on rating system design (e.g. [2][3]), our results show that a social strategy that recommends all users cooperate (i.e. to produce content) is often not optimal in the considered UGC networks where users are heterogeneous and when the monitoring is imperfect. We determine the condition on the monitoring accuracy when such an intuitive strategy is optimal and then develop a low-complexity algorithm that finds the optimal strategy within finite iterations when these conditions are not satisfied. Our results suggest that, surprisingly, allowing a certain level of freeriding behavior may lead to higher social welfare than incentivizing all users to produce. (Proofs available at [8].)

2. RELATED WORKS

The freeriding behavior of users has been observed in many types of UGC networks (e.g. [1]). Numerous existing incentive mechanisms rely on pricing mechanisms or reputation mechanisms to incentivize cooperation in networks. Pricing mechanisms are appropriate in many settings, but they are not adequate incentive schemes for UGC networks where much of the appeal is that the content is free. Much of the existing work on reputation mechanisms focuses on effective information gathering techniques (e.g. [4]) and empirical studies (e.g. [5]). The few works providing theoretical results consider either one (or a few) long-lived seller(s) interacting with many short-lived buyers [2] or anonymous users interacting in a random matching scenario [3]. None of these works consider the design of rating systems with heterogeneous users interacting over a (given) topology.

There is a big economics literature that studies sustaining cooperation among agents through reputation mechanisms. The seminar works (e.g. [6]) study social norms in a random matching setting and show that cooperation can be sustained via threats of contagions of bad behavior. Cooperation in settings where agent interact on networks is studied. The most related work probably is [7]. This work shows heterogeneity may preclude cooperative behavior as agents' preferences in terms of discounting can be very diverse. Therefore, partitioning a group into more homogeneous subgroups can enable cooperative behavior which might not be feasible otherwise. However, as opposed to our paper, they assume that agents are a prior identical when the neighborhood planner is choosing a neighborhood design. Moreover, the planner has much more power than ours - it determines which agent can interact with which agent. In our setting, agents are connected over an exogenously determined topology and therefore, the administrator's design is subjected to the topology.

3. SYSTEM MODEL

3.1 The Content Production Game

We consider a set $\mathcal{N} = \{1, 2, ..., N\}$ of users in a network connected according to a directed topology matrix $\mathcal{G}$ with $g_{i,j} = 1$ if there is a directed link from user i to user j and $g_{i,j} = 0$ otherwise. Note our model is general enough to include the undirected connection relationship of users. We assume that this underlying topology is predetermined for the following analysis and we do not consider the network formation process. Time is discrete. In each period, users can decide whether or not to produce content which is valuable to users who connect to them. Denote the action space of the user by $\mathcal{A} = \{0, 1\}$ where $a = 1$ stands for "produce" and $a = 0$ stands for "freeride" (not produce). Denote $b_{i,j} \in \mathbb{R}$ as the benefit that user i obtains by receiving the content produced by user j. Furthermore, we denote $b_i^{in}(\mathcal{N}) = \sum_{j:g_{i,j}=1} b_{i,j}$ and $b_i^{out}(\mathcal{N}) = \sum_{j:g_{i,j}=1} b_{j,i}$. Therefore $b_i^{in}(\mathcal{N})$ is the maximum aggregate benefit that user i can obtain from its connected users' content production and $b_i^{out}(\mathcal{N})$ is the maximum aggregate benefit that the connected users can obtain from user i's content production. Producing content is costly and we denote $c_i > 0$ as the cost for user i to produce content. We assume that freeriding incurs no cost and $c_i < \max\{b_i^{in}(\mathcal{N}), b_i^{out}(\mathcal{N})\}, \forall i \in \mathcal{N}$. This assumption states that that the cost for user i to produce content is smaller than (1) the benefit that it can possibly obtain from its connected users if all of them produce content and (2) the benefit that it can provide to users that connect with it if user i produces content. It indicates that the socially optimal actions are that all users produce content all the time (if there is no incentives involved). However, since producing only incurs a cost but no direct benefit without an incentive mechanism being deployed, all users will choose to freeride to maximize their own utilities. Finally, since users are long-lived in the network, we assume that users discount the next period benefits and costs with a discount rate $\delta \in (0, 1]$.

3.2 The Rating System

The objective of the network administrator is to design incentive mechanisms to provide users with incentives to produce content in order to maximize the social welfare (defined as the sum of average utilities of all users in the network). The "first-best" social welfare is achieved when all users produce content, i.e. $U^{first-best} = \sum_{i \in \mathcal{N}} (b_i^{in}(\mathcal{N}) - c_i)$.

In this paper, we design a simple but practical rating system aiming to maximize the social welfare of the UGC network. The rating system has a binary rating set $\Theta = \{0, 1\}$ where $\theta = 0$ (1) represents the low (high) rating. Users with different ratings have different access levels to others' produced content. User i with a rating θ is able to access its connected users' produced content with probability $p_{\theta,i}$ which is designed and implemented by the network administrator. In general, there are two cases regarding the constraints on these probabilities: (1) discriminative policies where $p_{\theta,i}$ may be different from $p_{\theta,j}$ for users $j \neq i$ even though they have the same rating and (2) non-discriminative policies where these access probabilities only depend on users' ratings. Due to the space limitation, this paper only presents the analysis for non-discriminative policies which we consider to be more common in practical implementation.

The administrator recommends different actions for different users. The recommendation is a mapping $\sigma : \mathcal{N} \to A^N$. Since A is a binary space, a recommendation partitions the user set into two complementary subsets $\mathcal{P}$ and $\bar{\mathcal{P}}$. Only users in $\mathcal{P}$ are recommended to produce content. Therefore, we conveniently write a recommendation as $\sigma_{\mathcal{P}}$. We call $\sigma_{\mathcal{N}}$ the "all users produce" recommendation and all the others "part of users produce" recommendations.

The rating update rule, which is executed at the end of each period, decides how the ratings of the users should be updated according to their content production actions. The rating update rule is a mapping $\phi : \Theta \times A \to [0, 1]$ where $\phi(\theta, a)$ is the probability that the next period rating is θ if the current production action is a. In particular, we consider the following simple update rule. Suppose the recommendation is $\sigma_{\mathcal{P}}$: For user $i \in \mathcal{P}$, $\phi(1, 1) = 1 - \epsilon$, $\phi(0, 1) = \epsilon$,

$\phi(1,0) = 0$, $\phi(0,0) = 1$; For user $i \in \bar{\mathcal{P}}$, $\phi(1,\cdot) = 1$, $\phi(0,\cdot) = 0$. In words, if a user is not recommended to produce, then its rating is constantly high. Therefore, it is always in their self-interest for these users to freeride. If a user is recommended to produce, then the rating update depends on the monitored production outcome. The user will receive a high rating if it follows the recommendation (i.e. produce) and will receive a low rating otherwise. However, since monitoring users' production behavior is imperfect, there is some error probability ϵ such that even though the user produced content, the monitoring result is negative and hence, its next period rating is updated to low.

4. OPTIMAL RATING SYSTEM DESIGN

4.1 Optimal Access Probabilities for a Given Recommendation

We first study the optimal rating system design for a given recommendation strategy $\sigma_{\mathcal{P}}$. Therefore, the only design parameters are two content access levels $p_{0,i} = p_0, p_{1,i} = p_1, \forall i$ in the non-discriminative policy case. Denote $b_i^{in}(\mathcal{P}) = \sum_{j:g_{i,j}=1, j\in\mathcal{P}} b_{i,j}$ and $b_i^{out}(\mathcal{P}) = \sum_{j:g_{i,j}=1, j\in\mathcal{P}} b_{j,i}$.

Since users are long-lived in the network, they care about their long-term utilities. For a user $i \in \mathcal{P}$ who complies with the recommendation, its long-term utility is as follows if all other users follow the recommendation,

$$U_i^\infty(\theta) = (p_\theta b_i^{in}(\mathcal{P}) - c_i) + \delta[(1-\epsilon)U_i^\infty(1) + \epsilon U_i^\infty(0)].$$

The long-term utility if the user unilaterally deviates from the recommendation is

$$\tilde{U}_i^\infty(\theta) = p_\theta b_i^{in}(\mathcal{P}) + \delta U_i^\infty(0).$$

Using the one-shot deviation principle in the repeated games theory, incentive-compatibility requires

$$U_i^\infty(1) - U_i^\infty(0) \geq \frac{c_i}{(1-\epsilon)\delta}, \forall i \in \mathcal{P}. \quad (1)$$

The optimal content access probabilities are those that achieve equality in (1). We define $\gamma(\mathcal{P}) = \min_{i\in\mathcal{P}} b_i^{in}(\mathcal{P})/c_i$ as the critical benefit-to-cost-ratio (BCR) for a subset $\mathcal{P}$.

THEOREM 1. *For a given $\sigma_{\mathcal{P}}$, an IC rating system can be constructed if and only if $\gamma(\mathcal{P}) \geq \frac{1}{(1-\epsilon)\delta}$ and the optimal content access probabilities are $p_1^* = 1, p_0^*(\mathcal{P}) = 1 - \frac{1}{(1-\epsilon)\delta\gamma(\mathcal{P})}$.*

Theorem 1 has two implications. First, in order to design an IC rating system, the BCR of the users should be larger than a threshold. Second, it is always optimal to provide high-rated users with the highest content access level (i.e. $p_1 = 1$). However, the optimal access level for low-rated users depends on the system parameters as well as the recommendation-dependent critical BCRs. The optimal value of p_0 in fact represents the tradeoff between incentive-compatibility and the social welfare. If $p_0 > p_0^*(\mathcal{P})$, then the rating system is not IC and there is at least one user who does not follow the recommendation. If $p_0 < p_0^*(\mathcal{P})$, then the social welfare is lower since users will obtain less benefit due to a lower access level to content when they drop to low ratings. Using the optimal content access levels, the optimal social welfare given $\sigma_{\mathcal{P}}$ is

$$U^*(\sigma_{\mathcal{P}}) = \left(1 - \frac{\epsilon}{(1-\epsilon)\delta\gamma(\mathcal{P})}\right)\sum_{i\in\mathcal{P}} b_i^{in}(\mathcal{P}) + \sum_{i\in\bar{\mathcal{P}}} b_i^{in}(\mathcal{P}) - \sum_{i\in\mathcal{P}} c_i$$

4.2 Optimality of the "All Users Produce" Recommendation

The recommendation strategy space is huge. For a network with N users, the cardinality of this space is 2^N. Intuitively, the optimal recommendation should be the "all users produce" recommendation. If this intuition was correct, then the administrator could simply recommend $\sigma_{\mathcal{N}}$ and design the corresponding optimal access levels. This can significantly reduce the complexity of solving the rating system design problem. Therefore, it is important to understand when this recommendation is indeed optimal.

THEOREM 2. *If the monitoring error probability is sufficiently small, i.e.*

$$\epsilon \leq \min\left\{1 - \frac{1}{\delta\gamma(\mathcal{N})}, \epsilon_\alpha = \frac{\alpha}{1+\alpha}\right\} \quad (2)$$

where

$$\alpha = \frac{\min_{i\in\mathcal{N}}(b_i^{out}(\mathcal{N}) - c_i)\delta\gamma(\mathcal{N})}{\sum_{i\in\mathcal{N}} b_i^{out}(\mathcal{N})} \quad (3)$$

then the "all users produce" recommendation is optimal.

The following corollary is immediately obtained.

COROLLARY 1. *As $\epsilon \to 0$, $U^*(\sigma_{\mathcal{N}}) \to U^{first-best}$.*

4.3 Optimal Recommendation

The question remains when the monitoring error is larger. In the following, we show that for a wide range of monitoring errors we can construct the optimal recommendation using a very simple method. In particular, we relax the error probability range to be $\epsilon_\alpha < \epsilon < \epsilon_\beta$, where $\epsilon_\beta = \frac{\beta}{1+\beta}$ and

$$\beta = \min_{\mathcal{P},\mathcal{Q}}\{\frac{\sum_{i\in\mathcal{P}-\mathcal{Q}}(b_i^{out}(\mathcal{N}) - c_i)}{\sum_{i\in\mathcal{P}} b_i^{in}(\mathcal{P}) - \sum_{i\in\mathcal{Q}} b_i^{in}(\mathcal{Q})}\delta\gamma(\mathcal{N}) \text{ where} \quad (4)$$

$$\mathcal{P} \subseteq \mathcal{N}, \mathcal{Q} \subseteq \mathcal{P} \text{ and } \sum_{i\in\mathcal{P}} b_i^{in}(\mathcal{P}) - \sum_{i\in\mathcal{Q}} b_i^{in}(\mathcal{Q}) > 0\}. \quad (5)$$

Note that (1) $\beta \geq \alpha$ and (2) β is much larger than α in most scenarios. Under this assumption, we establish the following important lemma.

LEMMA 1. *Suppose $\epsilon \leq \epsilon_\beta$. If two recommendations $\sigma_{\mathcal{P}}$ and $\sigma_{\mathcal{Q}}$ satisfy $\mathcal{Q} \subset \mathcal{P}$, $\gamma(\mathcal{P}) \geq \gamma(\mathcal{Q})$ and $\gamma(\mathcal{P}) \geq \gamma(\mathcal{N})$, then the social welfare satisfies $U^*(\sigma_{\mathcal{P}}) \geq U^*(\sigma_{\mathcal{Q}})$.*

The above lemma indicates that a recommendation $\sigma_{\mathcal{P}}$ is better (i.e. it leads to higher social welfare) than another recommendation $\sigma_{\mathcal{Q}}$, if certain simple conditions hold. Note that the $\mathcal{P} = \mathcal{N}$ satisfies $\gamma(\mathcal{P}) \geq \gamma(\mathcal{N})$ and therefore, we have the following theorem.

THEOREM 3. *Suppose $\epsilon \leq \min\{1 - \frac{1}{\delta\gamma(\mathcal{N})}, \epsilon_\beta\}$. If any "part of users produce" recommendation that has a lower critical BCR than the "all users produce" recommendation, then the "all users produce" recommendation is optimal.*

Theorem 3 is theoretically important but not very useful in practice. For one reason, applying Theorem 3 involves solving the BCRs for all possible recommendation strategies. This requires almost the same computational complexity as performing an exhaustive search. Moreover, if the condition

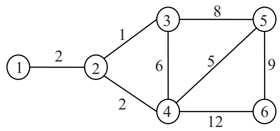

Figure 2: An Example Network.

in Theorem 3 does not hold, we are still not able to know which recommendation is optimal. Fortunately, using Lemma, we can construct a simple and efficient algorithm to find the optimal recommendation. We call it the "Iterative Deletion (ID)" algorithm and present it in Algorithm 1.

Algorithm 1: Iterative Deletion (ID) Algorithm

Input: $\mathcal{N}$, $\mathcal{G}$, $b_{i,j}, \forall i, j$ and $c_i, \forall i$.
Output: Optimal σ^* and p_0^*.
Set $\mathcal{P} = \mathcal{N}$
repeat
  Compute $\gamma(\mathcal{P})$
  if $\gamma(\mathcal{P}) > \gamma(\mathcal{N})$
    Compute $U^*(\sigma_{\mathcal{P}})$
  end
  Find $i := \arg\min_{i \in \mathcal{P}} b_i^{in}(\mathcal{P})/c_i$
  Update $\mathcal{P}$ by deleting i from $\mathcal{P}$
until $\mathcal{P} = \emptyset$
The optimal σ^* and p_0^* are those associated with the maximum $U^*(\sigma_{\mathcal{P}})$ and $\gamma(\mathcal{P}^*) \geq \frac{1}{(1-\epsilon)\delta}$ in the above iterative process.

The key idea of this algorithm is that in each iteration, we construct a subset $\mathcal{P}$ (so the recommendation is $\sigma_{\mathcal{P}}$) by deleting the user who has the critical BCR from the subset obtained in the last iteration. In this way, we hope to use a higher access level p_0 for low-rated users. The optimal recommendation is proven (in Theorem 4) to be among those emerging on the iterative path.

THEOREM 4. *Suppose $\epsilon \leq \epsilon_\beta$. The ID algorithm finds the optimal recommendation in exactly N iterations.*

5. ILLUSTRATIVE EXAMPLES

In this section, we provide an example to illustrate the importance of determining the optimal recommendation and how the ID algorithm works. Consider a network with $N = 6$ users on the topology shown in Fig. 2. For illustrative purpose, we assume that the costs for all users are the same $c_i = 1, \forall i \in \mathcal{N}$ and the benefits between two connected users are symmetric, i.e. $b_{i,j} = b_{j,i}$. The exact values of the benefits are shown on the edges in the figure. The discount factor is taken to be $\delta = 0.9$. Thus, we can compute $\epsilon_\alpha = 0.02$ and $\epsilon_\beta = 0.31$. The first-best social welfare is $U^{first-best} = 84$. We perform the ID algorithm for monitoring error probability being $\epsilon = 0.1$ and $\epsilon = 0.3$.

The partitions emerging on the algorithm path and their corresponding optimal social welfare for $\epsilon = 0.1$ and $\epsilon = 0.3$ are reported in Table 1 and Table 2, respectively. The partitions emerging on the algorithm path are the same for both cases since the deletion process does not depend on ϵ. However, the optimal social welfare is different and hence, the optimal recommendations are also different. For $\epsilon = 0.1$, the optimal recommendation is $\mathcal{P} = \{2, 3, 4, 5, 6\}$. For $\epsilon = 0.3$, the optimal recommendation is $\mathcal{P} = \{3, 4, 5, 6\}$. As we see, when the monitoring error probability is large, the optimal

t	User 1	User 2	User 3	User 4	User 5	User 6	$U^*(\sigma_{\mathcal{P}})$	IC
1	**2**	5	15	25	22	21	78.44	Yes
2	--	**3**	15	25	22	21	**79.46**	Yes
3	--	--	**14**	23	22	21	78.29	Yes
4	--	--	--	17	**14**	21	64.54	Yes
5	--	--	--	**12**	--	12	43.75	Yes
6	--	--	--	--	--	**0**	0	-

Table 1: Partitions and social welfare emerging on the algorithm path for $\epsilon = 0.1$.

t	User 1	User 2	User 3	User 4	User 5	User 6	$U^*(\sigma_{\mathcal{P}})$	IC
1	**2**	5	15	25	22	21	62.57	Yes
2	--	**3**	15	25	22	21	69.35	Yes
3	--	--	**14**	23	22	21	**76.28**	Yes
4	--	--	--	17	**14**	21	63.23	Yes
5	--	--	--	**12**	--	12	43.04	Yes
6	--	--	--	--	--	**0**	0	-

Table 2: Partitions and social welfare emerging on the algorithm path for $\epsilon = 0.3$.

"part of users produce" recommendation significantly outperforms the simple "all users produce" recommendation (22% improvement when $\epsilon = 0.3$ in this example). Therefore, it is of great importance for the network administrator to determine the optimal recommended strategy according to the accuracy of the monitoring technology.

6. CONCLUSIONS

In this paper, we studied how to design rating systems aimed at maximizing the social welfare of UGC networks. We showed that it is possible to exploit the ongoing nature of users' interactions to design rating systems to incentivize users to actively participate in content production. Our analysis showed that the imperfect monitoring and the user's heterogeneity in terms of both their content valuation and specific connectivity strongly influence the users' self-interested decisions and incentives. Surprisingly, in some scenarios, allowing a certain level of freeriding behavior in the UGC networks can achieve higher social welfare than incentivizing all users to produce content. This significantly differs from existing works in which incentivizing all users to cooperate is always optimal.

7. REFERENCES

[1] A. Kittur, B. Shu, B. A. Pendleton and E. H. Chi. 2007. He says, she says: conflict and coordination in Wikipedia. SIGCHI, pp. 453-462.
[2] C. Dellarocas. 2005. Reputation mechanism design in online trading environments with pure moral hazard. Information Systems Research, 16(2):209-230.
[3] Y. Zhang, J. Park, M. van der Schaar. 2013. Rating protocols for online communities. ACM Transactions on Economics and Computation (forthcoming).
[4] S. Kamvar, M. T. Schlosser, and H. Garcia-Molina. 2003. The eigentrust algorithm for reputation management in p2p networks. WWW 2003.
[5] P. Resnick and R. Zeckhauser. 2002. Trust among strangers in internet transactions: empirical analysis of eBay's reputation system: 127-157.
[6] M. Kandori. 1992. Social norms and community enforcement. Review of Economic Studies, 59: 63-80.
[7] M. Haag and R. Lagunoff. 2006. Social norms, local interaction, and neighborhood planning. International Economic Review. 47.1: 265-296.
[8] omitted proof, available online http://ee.ucla.edu/~jiexu/documents/UGC_tp.pdf