Event Estimation Accuracy of Social Sensing with Facebook for Social Internet of Vehicles

Kardelen Cepni, Mustafa Ozger, *Member, IEEE,* and Ozgur B. Akan, *Fellow, IEEE*

Abstract—Social Internet of Vehicles (SIoV) is a new paradigm that enables social relationships among vehicles via the Internet. People in the vehicles using Online Social Networks (OSNs) can be an integral part of SIoV that enables the collection of data for sensing a physical phenomenon, i.e., social sensing. In this paper, we study the main social sensing mechanism in Facebook, Comment Thread Network (CTN), which is based on the interactions of users through user walls in Facebook for SIoV. After seeing their commuters' contents about an event, users either add comments or like these posts, and Facebook CTN emerges as a social sensing medium in estimation of an event through social consensus. For the first time, this study investigates the social sensing capability of Facebook CTN, i.e., the accuracy of collective observations for SIoV. The accuracy depends on the user characteristics and the features of the OSN, since perceptions of the users and how they use Facebook may manipulate their observation signals. We analyze the reliability of Facebook CTN for varying user behaviors, user relationships, Facebook features, and network size. The results indicate that the polarized weighting of the observations and the use of less reliable post types in CTN deteriorate the accuracy of the estimate signal, i.e., social consensus. Furthermore, the selection of users is likely to be an important factor in social sensing.

Index Terms—Event Estimation, Social Sensing, Social Networks, Facebook, Social Internet of Vehicles.

I. INTRODUCTION

Social Internet of Vehicles (SIoV) provides connectivity between vehicles via the Internet and forms a network that facilitate collective and interactive communication between them for road safety and traffic management [1], [2]. Although SIoV envisions that vehicles form a machine-to-machine vehicular social networks using vehicular ad hoc networks architectures [4], people can be part of SIoV by behaving as social sensors [3]. With SIoV concept with commuters and passengers, they can provide inputs for extracting event information from their observations using online social networks (OSNs) such as Facebook. Our problem domain in this paper is estimation of events on the road from observations of the people in vehicles through social networks.

Social sensing, utilization of the information shared in OSNs for sensing a physical phenomenon, is a recently new research field of interest in computer science and associated disciplines. Increased use of OSNs enabled the collection and analysis of data from OSNs (including Twitter, Instagram and Facebook) to understand collective observations on an event of interest.

The authors are with the Department of Electrical and Electronics Engineering, Koc University, Istanbul, 34450 Turkey, e-mail: {kcepni, mozger, akan}@ku.edu.tr.

Ozgur B. Akan is also with Electrical Engineering Division, Department of Engineering, University of Cambridge, UK, e-mail: oba21@cam.ac.uk.

The significance of the collective observations appears in major social events and online marketing strategies. During major social events, people use OSNs to learn about the general situation regarding the event [5], [6]. As an example, the active use of OSNs during 2011 Arab Spring and 2013 Boston Marathon Bombings illustrates the significance of the information dissemination through OSNs [7], [8]. When the global information sources were not available for information retrieval, people followed the course of events by following users' posts in OSNs. Furthermore, fast information spreading capability of OSNs has provided firms to market their products through Online Marketing strategies [9]. The strategies include maintaining a product page in an OSN, in which they share advertorial information about the product, and the users share their opinions about the product. The observations collected in these pages may be an important signal indicating the performance of the product on the market. Therefore, understanding the social sensing mechanisms and the accuracy (or reliability) of collective observations in OSNs is an important problem.

Social sensing can also be used for different applications. For instance, social perception is studied by creating Wi-Fi signal based system to detect the presence of others for people who are blind or have social disabilities [10]. Furthermore, identification of interesting places in a city can be performed via social sensing applications by collective contribution by social sources such as humans and their smart phones [11]. Furthermore, people-centric or social sensing can be utilized in the Internet of Things and Internet of Vehicles [4] domain to complement physical sensing. In this domain, the authors in [12] investigate the ways how to integrate physical sensing with the social sensing to enable smart city mobility services. The services and applications offered by social sensing will increase.

Up to now, the research on social sensing has mainly evolved in two directions. First, there are studies that perform event detection for various events with OSNs (mostly Twitter) by sentiment analysis on the collected data [13]–[16]. This approach removes the uncertainty in reliability of social sensing by eliminating false or irrelevant user data, however, we are interested in the uncertainty of user observations since social sensing may be required during instantaneous situations, in which the analysis of data is not feasible and the decision-making authorities (e.g., governments in the case of disasters) may rely on the collective observations. In the second approach, several studies introduce the subject from a more abstract level, in which they define OSN users as *social sensors*, to understand the *social* and *uncertain* nature of social sensing. However, they do not provide theoretical

understanding of the process on how user observations are formed and how the consensus is reached [17]–[19]. Differently, [20] introduces social sensing from a theoretical level by utilizing social learning and game-theoretic models to investigate the interactions between the users by focusing on decision-making applications. However, the study does not consider the effect of OSN in use during social sensing. In social sensing, OSNs provide the media for generating user observations, hence studying the distinct mechanisms of OSNs is essential in understanding the reliability of social sensing. In this study, we direct our attention to Facebook, which is the most popular OSN (with respect to statistics in [21]), yet whose social sensing capabilities have not been studied from a theoretical level before.

Facebook, currently with 1.8 billion active users, is a popular free OSN that allows registered users to create profiles, upload photos and videos, send messages and keep in touch with friends, family and colleagues [22]. There exists several studies in literature, which investigate the network characteristics and user activities in Facebook by analyzing collected data [9], [23], [24]. On the other hand, [26] studies the effect of social influence of a Facebook fan page on movie box offices and develops an influence model to forecast box offices for movies, which shows the potential of Facebook as a medium for social sensing.

Facebook has indeed several features that provide a medium for social sensing. For example, *Opinion Polls* are the platforms that allow users to survey the opinions of its friends. Similar to Twitter, Facebook also has *Hashtag Pages*, in which Facebook posts labeled with a specific keyword are collected in a single Facebook page, and the available data may be utilized for social sensing. More importantly, the social sensing mechanism that is most authentic to Facebook is based on the interactions of users through user walls, the Facebook pages in which users either add comments or like the posts that are shared by their friends [23]. The discussion of the opinions through these interactions on Facebook continually yields in a collection of user posts, which is called as *comment thread* [25]. Comment thread is a medium for social sensing since it facilitates the collection of the opinions of Facebook users on a specific event, which may be estimated using the posts in the comment thread. We name the overall opinion formation process as *Comment Thread Network* (CTN), and in this study, we investigate the social sensing capabilities of Facebook CTN for SIoV, which, to the best of our knowledge, have not been studied so far.

Social sensing with CTN is radically different from the existing sensing schemes due to several factors, which stem mainly from the *social sensors*, i.e., Facebook users along with the features of Facebook. Observation signals are generated by Facebook users through CTN, and they include noise component due to users' perceptions. Moreover, sharing observation signals in various forms in the comment thread may introduce additional noise. After all, studying social sensing mechanisms and the social sensors in OSNs will foster the research on OSNs, and the gained insights are likely to contribute the effective designs of future OSNs along with social sensing applications in many areas in our lives mainly including

disaster management, online marketing and many more to come in future. We investigate its application on SIoV domain.

In this study, we develop an analytical social sensing model for the formation of user observations and social consensus in Facebook CTN as a part of SIoV from a signal processing perspective, by modeling user observations in vehicles as *social signals*, and we investigate the accuracy of collective observations for varying user behaviors, relationships and features of Facebook in SIoV. For the notion of social sensing contains concepts from statistical signal processing along with social learning models in economics, we will use the following group of words *estimate signal-social consensus-collective observations* and *user post-observation signal* interchangeably throughout the study.

The remainder of the paper is organized as follows. In Section II, we introduce the concept of Facebook CTN, describe the main features of Facebook in social sensing and state our assumptions on the Facebook CTN model. In Section III, we develop an analytical social sensing model and derive the estimation error function. In Section IV, numerical analysis is performed and the results are presented. Finally, we conclude the paper in Section V.

II. PRELIMINARIES ON FACEBOOK COMMENT THREAD NETWORK

Social sensing with CTN in Facebook is a sequential process, in which user observations evolve in time, and most comments in comment thread are made in few hours [29]. Therefore, social sensing with CTN may be studied in a limited time window. In a discrete time setting, we limit the time window with the number of participating users in CTN assuming that given an initial post, a single comment is generated at each discrete time step. Briefly, CTN is described as follows. Denoting the present time by n and assuming that the social sensing process ends at time n with N commenting users along with the initial user, who started the comment thread, we examine the course of user actions in CTN from past to present. The illustration of the process is shown in Fig. 1. Starting from the initialization of the comment thread, the initial Facebook user U_0 creates a post $p_0(n-N)$ at time $n-N$, which includes its observation about an ongoing event θ. Due to features of Facebook, this post appears in the Facebook pages of the users who are friends with U_0. We may call this group of the people as the *community C* of U_0. After seeing the post, the users in C may want to share their observations on the initial post. At the next time step, first commenting user, U_1 comments on the initial post with $p_{i+1}(n-N+1)$. Similarly, new posts are created sequentially (by the users $U_2, U_3...U_N$), and at time n, we obtain the set of posts (or comments) P

$$P = \{p_0(n-N), p_1(n+1-N), ..., p_N(n)\} \qquad (1)$$

Besides the set of posts P, Facebook releases an additional information about the popularity of the comments, which is the like count information of each post, denoted by $L(p_i)$ or simply L_i. Some users in C may also share their attitudes about the event by liking the relevant user posts in the
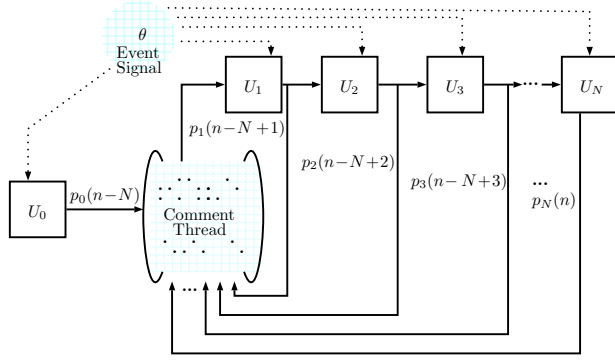
Fig. 1. Comment Thread Network Model

comment thread. The final set of like count information of the posts, L becomes

$$L = \{L_0, L_1, ..., L_N\}. \tag{2}$$

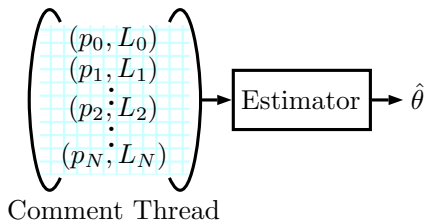Ultimately, using the available data in the comment thread

Comment Thread

Fig. 2. Social Consensus Formation in CTN

(i.e., the sets P and L) provided by the community, an estimation $\hat{\theta}$ on θ is made as shown in Fig. 2, in which the estimator block represents the social consensus. In the next section, we briefly describe and assess the main user action related features of Facebook in social sensing with CTN, i.e., comment and like actions, in order to comprehend Facebook CTN completely.

A. User Actions in Facebook CTN

In Facebook CTN, users generate the observation signals through several ways, i.e., by *commenting* and/or *liking* a given Facebook post.

1) Comment: By definition, a *comment* is a remark that expresses a personal reaction or attitude about a given fact. In Facebook, given a user's post, the friends of the user may share their opinions about the post via comment, a feature which allows users to add their opinions under the user's post. The opinion reflects their attitudes and recently learned facts about the state of event. Comments (comment posts) in Facebook may include various types of data such as text, image, video or a link to a web page.

2) Like: Like is a Facebook feature, available for every user post, which allows other users to demonstrate their support on the shared opinion by clicking like button. In other words, it counts the number of likes that a post gets after it is created. The posts belonging to influential users and containing

valuable information are likely to get more likes than the other posts, and users on Facebook use like button for highlighting the post among the other posts [9]. Therefore, like count information may be important in defining the estimate signal $\hat{\theta}$ in Facebook CTN.

Before giving the detailed description of Facebook CTN model in Section III, in the next section, we give the assumptions on the user behaviors and CTN model in Facebook.

B. Model Assumptions

Social sensing with Facebook CTN is a process that evolves in a user's community through comments and likes generated by the user's friends. It is sequential in the sense that users' posts will be affected from preceding users' posts. In our model, we use a discrete time framework and assume that users post their comments in sequence at fixed time steps. It is also assumed that the user will make a comment based on its current belief and the observation shared by the preceding user. In CTN, we assume that each user in the community is a rational user in the sense that it is aware of the event to be estimated θ, hence it does not lie or share irrelevant observation.

III. COMMENT THREAD NETWORK MODEL

Social sensing with Facebook CTN mainly relies on the event of interest θ to be estimated and a community C, in which users' observations on θ are generated. We start our modeling with the user relationships in Facebook CTN.

Assume that the friendship relationships within the community C (of size $N+1$) are extracted from the complete social network as a weighted graph, which consists of a set of nodes (Facebook users) and an adjacency matrix W of size $(N+1)\times(N+1)$, whose elements w_{ij} are non-binary values between $[0, 1]$ representing the level of the friendship between U_i and U_j. For higher values of w_{ij}, users U_i and U_j have a closer relationship. In our modeling, we assume that the relationships are symmetric ($w_{ij} = w_{ji}$), and the values of w_{ij} for $i = j$ are ignored due to their irrelevance in defining friendship relationships.

In general, discussions on social issues such as politics, sport events, natural disasters and product reviews are seen in Facebook comment threads [27]. Therefore, statistical model of the phenomenon θ will heavily depend on the event of interest. For example, if the discussion is on the occurrence of an event (e.g., result of a soccer game), then the event may be modeled as a binary signal. On the other hand, the event of interest may not be about a polarized subject, yet it may be on a subject, in which user observations consist of a continuous range of standpoints. For example, in the estimation of weather temperature, user observations consist of scalar temperature values within a continuous range. In our model, we will simply take phenomenon to be an unknown constant scalar parameter, θ. In the next section, we introduce the social signal model for user observations in Facebook CTN.

A. User Observations

Main components of social sensing with Facebook CTN are the participating users, who act as sensors in observing and sampling the event of interest. They are social sensors, which perform sensing through their perceptions based on the sensory inputs from outside world. However, the users may introduce noise at different levels of sensing process arising from their perceptions of reality and the methods which they reflect their observations in comment thread. Thus, in this section, we decompose user observation signal into two signals, which are pure observation signal and observation encoding noise, and assess their characteristics.

1) Pure Observation Signal: A user's pure observation on the event θ is the reflection of event in user's mind through its perception, which may be dependent on the user including its sensory reception, personal characteristics and social status. This fact implies that U_i's observation may not be exactly θ due to variance in its perception and may change with time. To capture the impact of user perceptions in CTN model, we model user observation signals as jointly Gaussian random variables $\theta_i[n]$ for $i = 0...N$ with mean θ and variance σ_θ^2. Since social sensing in CTN occurs within a community and users are connected via friendship relationships, it is highly possible that users observations are correlated based on their friendship. Friendship distance between two users U_i and U_j is defined by

$$d_{i,j} = 1 - w_{ij}, \tag{3}$$

the higher the strength of friendship, the distance between two users will be less, hence their ideas will be more correlated. Considering the correlation resulting from time distance between users' posts and friendship relationship among the users, we assume a stochastic relationship between user observations of U_i and U_j. In a more explicit form, the correlation relationship between $\theta_i[x]$ and $\theta_j[y]$ (at times x and y, respectively) is given as

$$E(\theta_i[x]\theta_j[y]) = cov(\theta_i[x], \theta_j[y]) + \theta^2$$
$$= \sigma_\theta^2 C(d_{i,j}, x - y) + \theta^2, \tag{4}$$

where $C(.)$ is the covariance function which defines the correlation relationships between user observations with respect to friendship and time distance among the posts. In this study, we use a general multivariate exponential covariance function in the form of

$$C(d_{i,j}, x - y) = e^{\frac{-(1 - w_{ij})}{\theta_F}} e^{-\frac{|x - y|}{\theta_T}}, \tag{5}$$

where θ_F and θ_T are correlation coefficients which determine how quickly covariance falls off with respect to friendship and time distance, respectively, and f is the sampling rate of the observed event [28]. Sampling rate is related to how frequently posts are created in Facebook CTN. For simplicity, we assume that posts are generated in each every minute in correspondence with the discrete time framework (the smallest unit of time is a minute), and f rate is set to one sample per one minute.

2) Observation Encoding Noise: When U_i shares its pure observation $\theta_i[n]$ via a post on Facebook, it has to determine how to encode (or represent) this observation by choosing a proper post type. Facebook allows users to share data in many forms including text, image, video, sound or web link. The post types may cause different representations of the pure observation signal, which may change the information quality of the post, i.e., users' perceptions and learning on given pure observation signal [30]. The reason is the fact that the representation of a user's view may be perceived differently by the other users, and the used post type may affect this perception. For example, while a user can alter the content of its observation shared in a text-based post more easily, since the perception of the content heavily depends on the user's writing style, the user's manipulation capabilities are limited in the posts shared as an image or a link since there is no extra room for the user to change these posts in Facebook. Consequently, receiving the pure observation signal in a text-based post is more prone to errors than receiving it in other post types due to both mechanisms of human learning and user characteristics, i.e., the form, which U_i uses to represent its observation in, may introduce additional noise for social sensing.

Thus, the use of different post types is likely to change the overall social consensus by altering the observation signals. We may incorporate the effect of different post types by introducing additional noise in users' observations based on the post type they use in CTN, and we will refer to this noise as *observation encoding noise*.

In our model, we limit the available post types on Facebook by only considering text and links to web pages, and we assume that text-based posts are less reliable than link-based posts (including both image and text). Therefore, while forming its comment, the user will introduce additional observation encoding noise on its pure observation signal. Consequently, user U_i's overall observation $s_i[n]$ becomes

$$s_i[n] = a_i v_i^T[n] + b_i v_i^L[n] + \theta_i[n], \tag{6}$$

where a_i and b_i are variables with binary values of $\{0, 1\}$, which indicate the use of either a text based or a link based post type by the user, respectively, and $a_i + b_i = 1$ for all $i = 0...N$. In Eq. (6), $v_i^T[n]$ and $v_i^L[n]$ are the random variables that model the observation encoding noise for each user U_i and for each post type, i.e., text-based and link-based posts, respectively. In our model, we assume that $v_i^T[n]$ and $v_i^L[n]$ are i.i.d Gaussian random variables with zero mean and variances σ_T^2 and σ_L^2, respectively, and we assume that $\sigma_T^2 \geq \sigma_L^2$.

B. Comment and Like Process

CTN starts with the first user U_0's post $p_0(n - N)$

$$p_0[n - N] = s_0[n - N]. \tag{7}$$

In the next time step, upcoming user U_1 sees $p_0[n-N]$ and posts its comment $p_1[n-N+1]$, in which it takes into account preceding user's post $p_0[n - N]$ and its observation signal $s_1[n-N+1]$. The user updates its observation with a weighted average social learning rule by allocating weights on each

opinion [31]. The weight to be allocated for the preceding post is determined by U_0's friendship level with U_0, i.e., w_{10}. Thus, U_1's post $p_1[n-N+1]$ becomes

$$p_1[n-N+1] = w_{10}p_0[n-N] + (1-w_{10})s_1[n-N+1]. \quad (8)$$

Note that since time indices of the posts can be obtained from the information on the sequence of the users in CTN, we may simplify the notation as $P = \{p_0, p_1, ..., p_N\}$, where $p_i = p_i[n-N+i]$, and similarly $s_i = s_i[n-N+i]$.

Using the simplified notation, general form of Eq. (8) for commenting users reduces to

$$p_{i+1} = w_{(i+1)i}p_i + (1-w_{(i+1)i})s_{i+1}, \text{ for } i = 0...N-1 \quad (9)$$

The generation of user posts continues sequentially, as described in Section II, and P is obtained.

In addition to user posts, like count information, which is available as a set $L = \{L_0, L_1, ..., L_N\}$ in Facebook should also be considered while modeling social sensing with CTN. Like is the count information of a Facebook post, which may be considered as a signal indicating the favorability of the opinion shared. Given a user post p_i, like count of p_i, L_i, depends on several factors. First, given two Facebook posts, the post which is posted earlier than the other one, is likely to get more likes since its probability of being seen by more users is higher than the other post. Namely, each post p_i in CTN will appear for different amount of time in the Facebook pages of the users. Second, it is known that the author of the post, who is a more favorable person in the community than the author of the other post, will get more likes. Thus, we may model the number of likes L_i, which post p_i receives, with a Binomial distribution, in which L_i is the number of successes (likes) in a sequence of X (duration of the post) independent like attempts by a single user, each of which results in success with the probability f_i. Assuming that K users attempt to like the post at each time step, we may approximate the total number of likes L_i by its mean

$$L_i = XKf_i, \quad (10)$$

where $X = N-i+1$ is the number of attempts (duration of the post), K is the number of Facebook users who attempt to like the post at the given discrete time step, and f_i is the probability that the post is liked by a user, approximated by the favorability (popularity) of the author of the post, U_k (for $k = 0...N$) as

$$f_k = \frac{\sum_{i=0, i\neq k}^{N} w_{ik}}{\sum_{i,j=0, i\neq j}^{N} w_{ij}}, \quad (11)$$

which is simply the normalized average of the weights given by the users in C to U_k. In other words, f_k reflects the user's relationship strength with the community. The higher values of f_k (compared to other users) implies that U_k is a more favorable user in the community.

C. Facebook Comment Thread Network Estimator

The impact of each user's observation on the estimate signal (social consensus) $\hat{\theta}$ in CTN depends on the like count information of the posts, since like counts in Facebook demonstrate the level of the support on the post and increase its visibility among the other posts. Given that N users post on p_0, user posts accumulate in comment thread, and ultimately we obtain the collection of data $\{(p_0, L_0), (p_1, L_1), ..., (p_N, L_N)\}$, the elements in each tuple represent the comment signal and like count information of the corresponding user, respectively. The general agreement $\hat{\theta}$ on θ may be obtained from the observations

$$\hat{\theta} = \frac{\sum_{i=0}^{N} L_i p_i}{\sum_{i=0}^{N} L_i}, \quad (12)$$

which is the weighted average of user observations with respect to their like counts. By explicitly writing the expression in Eq. (12) using Eq. (9) and Eq. (10), we obtain

$$\hat{\theta} = \frac{\sum_{i=0}^{N} \left(L_i(1-w_{(i+1)i}) + L_{i+1}w_{(i+1)i}\right)s_i}{\sum_{i=0}^{N}(N-i+1)Kf_i}. \quad (13)$$

Since we assume that w_{ij}, L_i, K and p_i are fixed parameters for all users in CTN, we simplify the expression in Eq. (13) as

$$\hat{\theta} = \frac{\sum_{i=0}^{N} c_i s_i}{M}, \quad (14)$$

where

$$c_i = L_i(1-w_{(i+1)i}) + L_{i+1}w_{(i+1)i}, \quad (15)$$

for $i = 0...N$ and $L_{N+1} = 0$, and

$$M = \sum_{i=0}^{N}(N-i+1)Kf_i. \quad (16)$$

In evaluation of the social sensing capabilities of Facebook CTN, we use the estimation error function while analyzing the accuracy of the estimates, defined by mean squared error (MSE)

$$D(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]. \quad (17)$$

The explicit form of $D(\hat{\theta})$ is given in Eq. (18).

IV. NUMERICAL ANALYSIS

In this section, we perform numerical analysis of the estimation error $D(\hat{\theta})$ (denoted as D in Figs. 3-7), using Eq. (18) for varying user actions and behaviors, Facebook features and network size in MATLAB. Friendship network on Facebook follows power law characteristics similar to general social network characteristics, however, this type of graph contains overall network structure including all users [24]. Since in this study we focus on small communities inside Facebook, we simply use random weighted graphs while modeling the relationships within the community and assume that all users in C are friends with each other. Furthermore, since $D(\hat{\theta})$

$$D(\hat{\theta}) = \theta^2 + \frac{1}{M^2}\sum_{i=0}^{N}\Big(a_i^2\sigma_T^2 + b_i^2\sigma_L^2 + \theta^2 + \sigma_\theta^2\Big)c_i^2 - \frac{2\theta^2}{M}\sum_{i=0}^{N}c_i + \frac{1}{M^2}\sum_{\substack{i,j=0 \\ i\neq j}}^{N}c_i c_j\big(\theta^2 + \sigma_\theta^2(e^{\frac{-(1-w_{ij})}{\theta_F}}e^{\frac{-|i-j|}{\theta_T}})\big) \qquad (18)$$

depends on the friendship relationships within the community defined by W, the analyses are iterated for various random weighted graphs (10000 times) and $D(\hat{\theta})$ is found by taking average of the results in these runs. Since we do not explicitly define θ in our modeling, $D(\hat{\theta})$ has no unit, and in the analyses, the default parameters in Table I are used unless otherwise stated.

TABLE I
NUMERICAL ANALYSIS PARAMETERS

Parameter	Value	Parameter	Value
N	20	θ_T	100
K	10	θ_F	1
θ	0	σ_T	0.5
σ_θ	0.2	σ_L	0.1
f	1 min^{-1}	a_i for $i=0...N$	1

The analysis results are presented in two sections. First, we investigate the changes in estimation accuracy for varying network size and post characteristics. In the second section, we mainly analyze the effect of user behaviors and relationships on social sensing with Facebook CTN.

A. Estimation Accuracy Analysis for Varying Network Size and Post Characteristics

The graph in Fig. 3 illustrates the changes in $D(\hat{\theta})$ for social sensing with CTN for varying event signal characteristics σ_θ along with varying number of users N. (Note that Facebook CTN has $N+1$ users in total, but in our analyses we represent it with the total number of commenting users N for simplicity.) In the analysis, the size of community (N) is taken within the range $(5, 50)$, which is the average number of comments in comment thread [25]. The analysis result shows that while for high values of event signal variance ($\sigma_\theta = 1$ and $\sigma_\theta = 5$), the estimator is defective and increasing N does not improve the accuracy of estimation, for relatively low values of signal variance ($\sigma_\theta = 0.2$), estimation accuracy is significantly higher and it improves with increasing number of participants. This trend indicates that CTN may not be suitable in social sensing for all event types, in which user observation signals are highly volatile.

In Fig. 4, the effect of variations in post type on estimation accuracy has been investigated. The x-axis values represent the percentage of text input data $(a_0+a_1+...+a_N)/(N+1)$ in all users' observations, which consist of text and link-based posts only. The figure depicts that as more text-based posts are included in CTN, the accuracy of social sensing deteriorates by approximately 5 times considering the link-based posts only. The reason is that while sharing observations in CTN, use

of data types which are more prone to inclusion of user bias (such as text) is likely to alter the consensus of social sensing adversely.

B. Estimation Accuracy Analysis for Varying User Behaviors and Relationships

While modeling *like* process, we assume that the most popular users get more likes compared to other users. In the analysis in Fig. 5, we compare the accuracy of social sensing for the case, in which we assume that all users are equally liked by their friends, rather than our initial assumption based on popularity-based liking. The graph in Fig. 5 indicates that the estimation is less erroneous with equal likes compared to popularity-based likes. Namely, average of all observation signals yields in better performance instead of amplifying the observations of the several users in the community. Furthermore, equally weighting of the observations makes social sensing more robust to more noisy data types.

By following a similar approach to the previous analysis, in the next analysis we investigate the effect of users' opinion weighting behavior while commenting in CTN. In the previous analyses, we assume that users weight preceding comments based on their friendship relationship with the corresponding user. However, sometimes users may choose to allocate more weight their own observations than the preceding user's comment in CTN. We call this type of behavior as *egocentric* commenting, in which users share only their observations by ignoring the preceding comments. In this scenario, $w_{i,i+1} = 1$ for all $i = 0...N$, however note that the observation correlations and like counts are still computed based on the friendship relationships. In the analysis, we also investigate the case
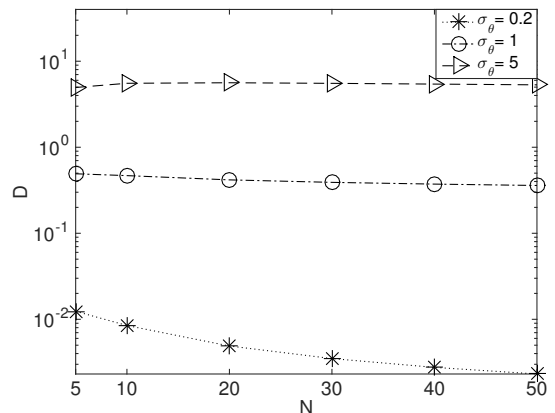

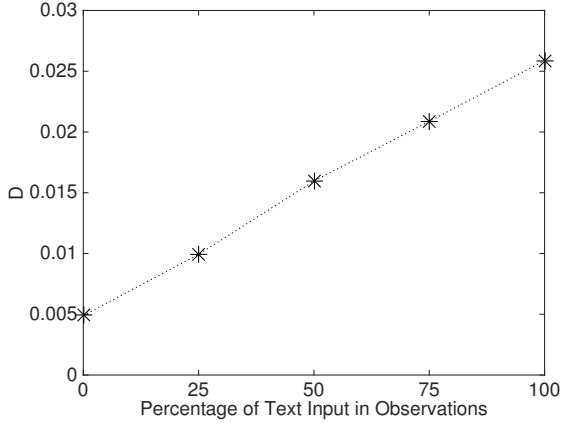Fig. 3. $D(\hat{\theta})$ for varying N and σ_θ

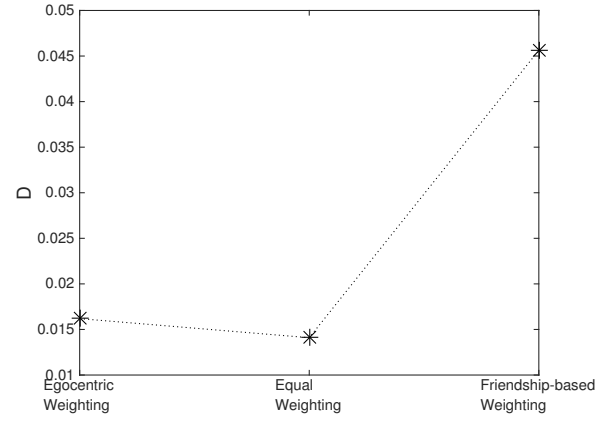Fig. 4. $D(\hat{\theta})$ for varying percentage of text-based posts in comment thread

Fig. 6. $D(\hat{\theta})$ for varying opinion weighting behaviors

in which the users behave temperately by equally weighting their observations against preceding comment. Fig. 6 shows the change in $D(\hat{\theta})$ for varying user behavior. As can be seen, friendship-based weighting results in estimation with a higher value of error (approximately 3 times order) compared to egocentric and equal weighting. These differences in $D(\hat{\theta})$ indicate that unequal weighting of observations via friendship-based weighting while forming social consensus is likely to result in more errors by focusing on few popular individuals' observations when compared to equal weighting.

In the last analysis, we investigate the effect of the correlation between the users in C, which we define using Eq. (5) for varying θ_F values that change the level of correlation between users' observations. The correlation relationship between users' observations is significant since the accuracy of social sensing in Facebook CTN may change depending on the characteristics of the community. This aspect may especially be important in social sensing applications regarding social events (e.g., political elections). In Fig. 7, the change in $D(\hat{\theta})$ is given for varying θ_F. In the analysis, we assume that all observations are equally weighted in order to investigate the effect of correlations more clearly. The analysis shows
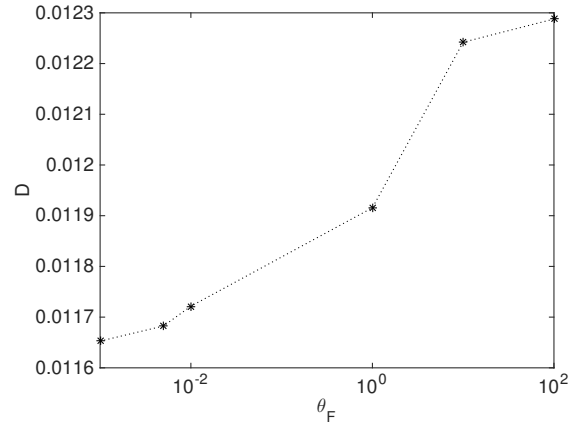

Fig. 7. $D(\hat{\theta})$ for varying θ_F

that estimation error increases by approximately 40% as the correlation between the users increases. In other words, less correlated observations benefit to the accuracy of social sensing, and increasing level of correlation in the community is likely to bias the social consensus. Therefore, the selection of community might be an important factor in social sensing with Facebook CTN.
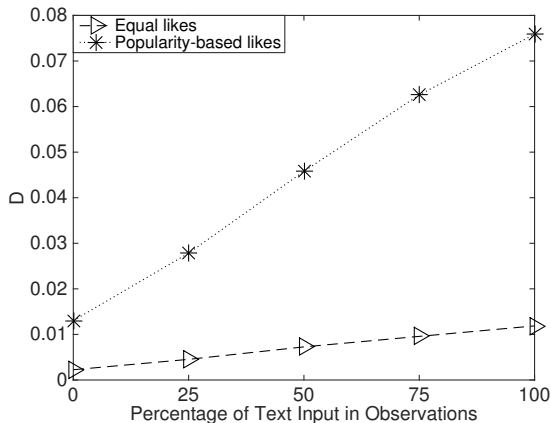
V. CONCLUSION

CTN is the main mechanism for social sensing in Facebook, which is actively used by Facebook users during the course of major social events for estimating an event of interest in SIoV. Understanding the reliability of estimation made out of collective user observations in CTN is significant due to applications of social sensing in numerous aspects of human lives. In this study, by developing an analytical social sensing model for user observations in CTN from a signal processing perspective, in which we model Facebook users as *social sensors* and their observations as *social signals* by incorporating their perceptions in SIoV, we analyze the reliability of Facebook CTN for varying user behaviors and relationships, Facebook features and network size. The results indicate that while the characteristics of the event signal affect the accuracy of the

Fig. 5. $D(\hat{\theta})$ for varying like behaviors

social sensing, also the polarized weighting of the observations and use of less reliable post types in CTN during social sensing are important in the sense that they deteriorate the accuracy of the estimate signal, i.e., social consensus. Furthermore, user relationships inside the community may alter the reliability of social sensing implying that the selection of OSN users in social sensing is essential in obtaining an accurate estimation. In our future research, we aim to investigate and understand the effect of friendship correlations between the users for various types of events and also the effect of liar users in Facebook CTN on the estimate signal.

REFERENCES

[1] K. M. Alam *et al.*, "Toward Social Internet of Vehicles: Concept, Architecture, and Applications," *IEEE Access,* vol. 3, pp. 343-357, 2015.

[2] L. A. Magralas *et al.*, "Social Internet of Vehicles for Smart Cities," *Journal of Sensor and Actuator Networks,* vol. 5, no. 3, pp. 1-22, 2016.

[3] K. Cepni, M. Ozger, and O. B. Akan, "Vehicular Social Sensor Networks," in *Vehicular Social Networks,* Boca Raton, FL: CRC Press, Taylor & Francis Group, 2017, pp. 19-34.

[4] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of Vehicles: From Intelligent Grid to Autonomous Cars and Vehicular Clouds," in *Proc. IEEE World Forum on Internet of Things,* 2014.

[5] H. Gao, G. Barbier, and R. Goolsby, "Harnessing the crowdsourcing power of social media for disaster relief," *IEEE Intelligent Systems*, vol. 26, no. 3, pp. 10–14, 2011.

[6] M. Greer and J. Ngo, "Personal emergency preparedness plan (pepp) facebook app: Using cloud computing, mobile technology, and social networking services to decompress traditional channels of communication during emergencies and disasters," in *Proc. ACM CHI'2011*, May 2011.

[7] A. Gupta, and H. Lamba, and P. Kumaraguru, "$1.00 per RT #Boston-Marathon #PrayForBoston: Analyzing fake content on Twitter," in *Proc. eCRS'2013*, 2013, pp. 1–12.

[8] A. Kavanaugh, S. Yang, S. Sheetz, L. Li, and E. Fox, "Microblogging in crisis situations: Mass protests in iran, tunisia, egypt," in *Proc. IEEE PASSAT'2011 and IEEE SocialCom'2011*.

[9] I. P. Cvijikj, E. D. Spiegler, and F. Michahelles, "The Effect of Post Type, Category and Posting Day on User Interaction Level on Facebook," in *Proc. IEEE PASSAT'2011 and IEEE SocialCom'2011*, Oct. 2011, pp. 810–813.

[10] Y. Halperin, G. Buchs, S. Maidenbaum, M. Amenou, and A. Amedi, "Social Sensing: a Wi-Fi based Social Sense for Perceiving the Surrounding People, in *Proc. ACM Augmented Human International Conference,* 2016.

[11] C. Huang, D. Wang, and B. Mann, "Towards social-aware interesting place finding in social sensing applications, *Knowledge-Based Systems (Elsevier)*, Feb. 2017, pp. 1-11.

[12] R. G. Qiu, L. Qiu, and Y. Badr, "Integrating Physical and Social Sensing to Enable Smart City Mobility Services, in *Proc. IEEE International Conference on Industrial Informatics,* 2016, pp. 909-915.

[13] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: Real-time event detection by social sensors," in *Proc. WWW'2010*, pp. 851–860, 2010.

[14] T. Sakaki, Y. Matsuo, T. Yanagihara, N. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors,"in *Proc. IEEE CYBER'2012*, pp. 221–226, May 2012.

[15] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. Ebert, and T. Ertl, "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," in *Proc. IEEE VAST'2012*, Oct 2012, pp. 143–152.

[16] R. Li, K. H. Lei, R. Khadiwala, and K.-C. Chang, "Tedas: A twitter-based event detection and analysis system,"in *Proc. ICDE'2012*, pp. 1273–1276, Apr. 2012.

[17] B. Guo, Z. Yu, D. Zhang, and X. Zhou, "Cross-community sensing and mining," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 144–152, Aug 2014.

[18] A. Rosi, M. Mamei, F. Zambonelli, S. Dobson, G. Stevenson, and J. Ye, "Social sensors and pervasive services: Approaches and perspectives," in *Proc. IEEE PERCOM Workshops'2011*, March 2011, pp. 525–530.

[19] A. Campbell, S. Eisenman, N. Lane, E. Miluzzo, R. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G.-S. Ahn, "The rise of people-centric sensing," *IEEE Internet Computing*, vol. 12, no. 4, pp. 12–21, July 2008.

[20] V. Krishnamurthy and H. Poor, "A tutorial on interactive sensing in social networks," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 1, pp. 3–21, March 2014.

[21] Duggan, M., Ellison, N.B., Lampe, C., Lenhart, A,. and Madden, M. "Social Media Update 2014," Pew Research Center, January 2015. Available at: http://www.pewinternet.org/2015/01/09/social-media-update-2014/

[22] Facebook Statistics. http://newsroom.fb.com/company-info/

[23] A. Muangngeon and S. Erjongmanee, "Analysis of facebook activity usage through network and human perspectives," in *Proc. IEEE KST'2015*, Jan. 2015, pp. 13–18.

[24] K. Nguyen and D. A. Tran, "An analysis of activities in Facebook," in *Proc. IEEE CCNC'2011*, Jan. 2011, pp. 388–392.

[25] L. Backstrom, Lars, J. Kleinberg, Jon, L. Lee, and C. Danescu-Niculescu-Mizil, "Characterizing and curating conversation threads: expansion, focus, volume, re-entry," in *Proc. ACM WSDM'2013*, Feb. 2013, pp. 13–22.

[26] W.-H. Tang, M.-Y. Yeh, and A. J. Lee, "Information diffusion among users on Facebook fan pages over time: Its impact on movie box office," in *Proc. IEEE DSAA'2014*, Oct. 2014, pp. 340–346.

[27] "Sandy is year's No. 2 topic on Facebook," http://edition.cnn.com/2012/10/30/tech/social-media/sandy-facebook-chatter/. Last accessed on April 5, 2015.

[28] *Linear Models for Multivariate, Time Series, and Spatial Data.* Springer-Verlag New York, 1991.

[29] "Local news in a digital age," Pew Research Center, Tech. Rep., 2015.

[30] "Multimodal Learning Through Media, Cisco Systems," Tech. Rep., 2008.

[31] C. Chamley, *Rational Herds: Economic Models of Social Learning.* Cambridge University Press, 2004.