Information in Tweets: Analysis of a Bufferless Timing Channel Model

Mehrnaz Tavan, Roy D. Yates, Waheed U. Bajwa
Department of Electrical and Computer Engineering
Rutgers University

***Abstract*—There has been a considerable interest in quantifying the influence that one node exerts on another in a social network. Using directed information, we study the problem for a simple, two-node network that models two users in a Twitter network in which one user (Alice) influences the other user (Bob) through her tweets. Under this setup, we relate the problem of direction of influence to the calculation of directed information from the input to the output in a bufferless single-server timing queue. Based on this relationship, we compute the directed information rate from Alice to Bob and, under simplifying assumptions, relate that rate to the distributions of Alice's tweet timings and Bob's action timings.**

I. Introduction

In recent years, there has been a growing interest in characterizing *influence graphs* in which a directed edge represents the influence that a node (actor) exerts on another [1]–[4]. Influence graphs are inferred from network interactions data, and are different from friendship graphs in which people are explicitly tied together as followers or friends [5]. Influence graphs have been used in the context of social networks such as Facebook and Twitter which are characterized by interactions among its actors (individuals, organizations, etc.), to determine the influential users [3], [4]. They also have applications outside the context of social networks. For example, [1] has used this approach to identify the underlying causal relationship graph in neural spike train recordings. Moreover, in mobile social network applications, influence graphs have been applied to routing packets among mobile users in ad hoc networks [6].

The diffusion of news items, tweets, and other types of information can be modeled as a cascading process over edges of an influence graph, analogous to a contagion that spreads an infection [7]. The time instances when nodes become infected by the contagion (e.g., tweet generation times, blog updating times, etc.) is a type of interaction data often used for inferring influence graphs [3], [4], [8]. The infection times create causality constraints such that one infected node can be under the influence of another only if the influential node was infected earlier in time. Furthermore, the influence is modeled as stronger when the difference in their infection times is smaller [3].

A model for causal influence was introduced in [9] as *Granger causality* where node A is under the influence of node B if past observation of B will reduce the mean square error of linear estimation of future behavior of A given past history of A [1]. Information-theoretic measures such as directed information [10], [11] and transfer entropy [12], although not addressing influence directly, generalize the Granger concept to causality for nonlinear models.

In this paper, we are interested in a mathematical understanding of the problem of inference of influence graphs in social networks. In contrast to earlier works, however, we take a bottom-up approach to the problem. Specifically, we focus on a simple, two-node social network that can be interpreted as a model for interactions on a Twitter network. We assume that the influence graph in this network corresponds to a single directed edge from one node, say, Alice, to the other node, say, Bob. In the context of Twitter, this setup corresponds to Alice tweeting at certain points in time and Bob reacting to some of these tweets and tweeting in return. In order to further simplify the analysis, we assume that all tweets from Alice and Bob are on the same topic. It follows in this simplified setting that Alice's influence on Bob is completely described through the timing information of Alice's and Bob's tweets. Our goal then is to interpret the timings of Alice's and Bob's tweets in terms of the influence that Alice exerts on Bob.

We approach this problem by relating the Twitter interactions of Alice and Bob to the dependence between the input and output of a timing channel that conveys information by the timing of consecutive packets rather than by their contents. In this paper, the influence is computed by characterizing the directed information between the input and output of *bufferless single-server timing queue* (SSTQ).

A bufferless SSTQ is a packet service facility in which incoming packets are discarded while a packet is already in service. The bufferless model is motivated by the observation that Twitter streams are inherently lossy. Since people have limited ability to process information, many tweets are ignored and only some tweets generate responses [13]. The bufferless queue is a simple model for a recipient Bob that ignores subsequent tweets while processing a previously received tweet. While this model is admittedly too simple to truly capture social interactions, it embeds a technical challenge in characterizing the directed information in timing data when there is no one-to-one correspondence between the tweet arrival and departure streams. In this paper, we address this challenge by characterizing directed information in a two-user Twitter interaction modeled by discrete-time timing channels described by bufferless SSTQs with iid service times.

The timing channel capacity of bufferless SSTQs was investigated in [14] in a continuous-time setting and in [15]

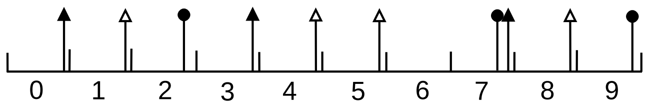

Fig. 1. One realization of the input and output sequence of the system is illustrated. The arrows with hollow arrowheads show the tweets arriving at the server that are dropped, the arrows with solid arrowheads show the tweets that enter the server, and the lines with circles on top show the tweets departing the server. The corresponding sequence of X_0^9 is $\{1,1,0,1,1,1,0,1,1,0\}$, and the corresponding sequence of Y_0^9 is $\{0,0,1,0,0,0,0,1,0,1\}$.

in discrete time. Since Twitter interactions between two nodes are unlikely to take place on infinitesimally small units of time, and more importantly since available timing data is quantized, we work with a discrete-time version of bufferless SSTQs in this paper. In contrast to continuous-time model in which directed information is accumulated by a receiver (an observer at the queue output stream) with each departure from the queue, a discrete-time model enables a finer-grained characterization of directed information. Specifically, with each discrete step in time, the departure or non-departure of a customer in service contributes to the directed information accumulated by the receiver.

The rest of the paper is organized as follows. In Section II, we provide an overview of our system, and a formal definition of directed information and directed information rate. Section III derives the directed information rate for M/G/1 queue. Concluding remarks are in Section IV.

Notation: We use $P_X(\cdot)$ to denote the probability mass function (PMF) of random variable X. Similarly $P_{X|Y}(\cdot|\cdot)$ is the conditional PMF of X given Y. For random variables, $\tilde{P}_X(x) = P[X > x | X > x-1]$. In addition, $h(\cdot)$ denotes the binary entropy function.

II. System Model

The key idea in the timing channels of [16] and [17] (the former in continuous-time and the latter in discrete-time) is to use packet inter-arrival times to the server to encode a message. The receiver, based on the departure times of packets from the server, decides which message has been transmitted. In this paper, we examine how models of this type can be used to evaluate the directed information associated with tweet streams where the server can model the thinking process of a person who tweets. Specifically, we consider the model in [15] consisting of a single server bufferless queue with a zero packet waiting room. Upon arrival at an idle server, a packet immediately enters service; otherwise, if the server is busy with a previous packet, the incoming packet is blocked and discarded. To be consistent with the definition of discrete-time queues in [17], we assume that in each time slot, at most one arrival and one departure can occur.

Returning to the Alice-Bob example of Twitter network, the sequence of tweets by Alice correspond to the packets that arrive at the server and the service time represents the thinking time for Bob in producing a tweet response. Note that tweets that arrive while Bob is thinking will be neglected and discarded. In this model, we assume that when a node tweets, the receiver node will see the incoming message immediately.

We use S_i to denote the service time of the ith tweet admitted to service. As is customary in discrete-time queues, we assume that service times are iid strictly positive integer-valued random variables, independent of tweet arrival times. We refer to the timing channel induced by the (bufferless) queue with service time S as the (bufferless) timing channel S. In Fig. 1, one realization of the input and output sequences, including arriving, dropped and departing tweets, is illustrated. In particular, a tweet that remains in service at the end of slot $i-1$ will receive one unit of service in slot i, and if its service is completed, depart the instant before the end of slot i. Furthermore, an arrival in slot i occurs at the end of the slot i, the instant after a possible departure. Thus, a tweet that arrives in slot i begins service in slot $i+1$ and departs no earlier than slot $i+1$.

To model the tweeting activity at the transmitter, the observation interval is divided into equal timeslots and the binary sequence $\overline{X} = (X_0 = 1, X_1, X_2, \cdots)$ will be associated with that process. In this system, $X_i = 1$ if a tweet is generated by Alice at the end of slot i. We refer to the tweet submitted at time 0 as tweet zero. This tweet carries no timing information and serves only to initialize the system.

At the receiver node, the sequence $\overline{Y} = Y_0, Y_1, \cdots$ corresponds to the discretized tweet process of Bob where $Y_i \in \{0,1\}$. Observing a departure (a tweet by Bob) at the end of slot i is represented by $Y_i = 1$.

The sequence of tweet inter-arrival times are represented by $\overline{A} = (A_0 = 0, A_1, A_2, \cdots)$ where $A_j \in \mathbb{N}$ is the inter-arrival time between tweets $j-1$ and j (in slots). The number of timeslots between departure $i-1$ and departure i correspond to D_i where D_0 is the departure time of tweet 0. In the bufferless queue, the subset of arrivals that are admitted into service is denoted by the subsequence $k_0 = 0, k_1, \cdots$ such that

$$k_i = \min\Big\{m | \sum_{j=1}^{m} A_j - \sum_{j=0}^{i-1} D_j \geq 0\Big\} \tag{1}$$

denotes the index of the tweet $i > 0$ admitted to service. The time that the server is idle between departure i and the next arrival is represented by the idling time W_i. Since the queue in our system is blocking and has no buffer, the idling time W_i can be represented as a deterministic function of the entire transmitted sequence and prior departures D_0^i as

$$W_i(A_0^\infty, D_0^i) = \sum_{j=1}^{k_{i+1}} A_j - \sum_{j=0}^{i} D_j. \tag{2}$$

Note that if the next arrival occurs in the same slot as the departure i, then $W_i = 0$. For ease of notation, we use $W_i(A_0^\infty, D_0^i)$ and the shorthand W_i interchangeably. The relationship between departure time D_i and the corresponding idling time and service time is

$$D_i = W_{i-1}(A_0^\infty, D_0^{i-1}) + S_i. \tag{3}$$

Since we assumed that in each time slot, at most one arrival and departure can happen, $A_i \geq 1$ and $D_i \geq 1$ for $i \geq 1$. Based on the assumption that in each timeslot, at most one arrival and one departure can happen such that if the next arrival occurs in the same timeslot the instant after a departure, the packet would enter the server with zero idling time so $W_i \geq 0$.

A. Directed Information

For discrete time stochastic processes X_1^n and Y_1^n, the directed information from X to Y is defined as [11]

$$I(X_1^n \to Y_1^n) = \sum_{i=1}^{n} I(X_1^i; Y_i | Y_1^{i-1}) = H(Y_1^n) - H(Y_1^n || X_1^n), \tag{4}$$

where $H(Y_1^n || X_1^n)$ is the causally conditioned entropy defined by Kramer [10] as

$$H(Y_1^n || X_1^n) = \sum_{i=1}^{n} H(Y_i | Y_1^{i-1}, X_1^i).$$

The relationship between the directed information and mutual information is

$$I(X_1^n \to Y_1^n) + I(DY_1^n \to X_1^n) = I(X_1^n; Y_1^n) \tag{5}$$

where $DX_1^n = (0, X_1, \cdots, X_{n-1})$ is the delay operator. Directed information is a nonnegative asymmetric quantity. When there is influence in both directions, the mutual information is an outerbound on the information that flows in each direction.

To analyze stochastic processes defined on an infinite time interval, we examine information rates. The directed information rate between two jointly stationary processes with finite alphabet is

$$I_\infty(X \to Y) = \lim_{n \to \infty} \frac{1}{n} I(X_1^n \to Y_1^n).$$

The proof for existence of this limit is in [10].

III. Analysis

With iid inter-arrival times, each time a tweet enters service, the queue undergoes a renewal. In particular, the ith renewal point marks the beginning of a service time S_i and a set of subsequent iid tweet inter-arrival times $A_{k_i+1}, A_{k_i+2}, \ldots$ such that the distributions of S_i and $\{A_{k_i+j}\}$ are sufficient to evaluate the distribution of the number of tweet arrivals that are dropped during the service as well as the idling time W_i that follows the service completion. We note that W_i depends on S_i; however the renewal that occurs when costumer $i+1$ goes into service implies that $(S_0, W_0), (S_1, W_1), \cdots, (S_n, W_n)$ constitute independent tuples.

The directed information rate from node A with tweet timing sequence X_0^∞ to node B with tweet timing sequence Y_0^∞ is

$$I_\infty(DX \to Y) = \lim_{n \to \infty} \frac{1}{n} \left[H(Y_0^n) - H(Y_0^n || DX_0^n) \right].$$

Before computing each term, we need the following Lemma.

Lemma 1. *For any nonnegative integer-valued random variable V, the following relationship holds*

$$H(V) = \sum_{l=0}^{\infty} h(\tilde{P}_V(l)) P[V > l-1]. \tag{6}$$

Proof: Denoting the right side of (6) as $\hat{H}$, the definition $\tilde{P}_V(l) = P[V > l | V > l-1]$ implies

$$\hat{H} = -\sum_{l=0}^{\infty} \left[P[V = l] \log\left(\frac{P[V = l]}{P[V > l-1]} \right) + P[V > l] \log\left(\frac{P[V > l]}{P[V > l-1]} \right) \right] \tag{7}$$

$$= H(V) - \sum_{l=0}^{\infty} \left[P[V > l] \log(P[V > l]) - P[V > l-1] \log(P[V > l-1]) \right], \tag{8}$$

which equals $H(V)$ since the sum in (8) is telescoping. ■

To characterize the bufferless G/G/1 queue, we observe that the tuple (A_i, B_i), where A_i is the age of the arrival process (i.e., the number of time units that have passed since the last arrival) and B_i is the units of service received by a customer in service (if any) at time i, is a discrete-time Markov chain. Note that $(0,0)$ denotes the state in which an arrival has just occurred but not yet received any service. In addition, we use $(A_i, B_i) = (a, -1)$ to denote a state in which the last arrival was in timeslot $i - a$ and the server is currently idle. We observe that state $(0,0)$ is reachable from any state (a, b) with a simultaneous departure and new arrival in the same slot. Because the system is bufferless, the limiting state probabilities $\pi_{a,b} = \lim_{i \to \infty} P_{A_i, B_i}(a, b)$ always exist. In addition, we can define the marginal limiting probabilities $\pi_b = \sum_a \pi_{a,b}$ that enable us to describe the bufferless G/G/1 queue.

Theorem 1. *For a bufferless stable G/G/1 SSTQ system,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} H(Y_i | Y_0^{i-1}, X_0^{i-1}) = \sum_{b=0}^{\infty} h(\tilde{P}_S(b+1)) \pi_b.$$

Proof: Since B_{i-1} is determined by (Y_0^{i-1}, X_0^{i-1}),

$$H(Y_i | Y_0^{i-1}, X_0^{i-1}) = H(Y_i | Y_0^{i-1}, X_0^{i-1}, B_{i-1}). \tag{9}$$

Furthermore $B_{i-1} = -1$ indicates that the server is idle and thus $Y_i = 0$ since there is no packet that can depart. Thus,

$$H(Y_i | Y_0^{i-1} = y_0^{i-1}, X_0^{i-1} = x_0^{i-1}, B_{i-1} = -1) = 0. \tag{10}$$

Otherwise, when a packet is in service and $B_{i-1} = b \geq 0$, the residual service time depends solely on the service b already received, and is independent of the prior arrival process and prior service times. Thus, since Y_i is a binary indicator for whether the current service finishes at time i,

$$H(Y_i | Y_0^{i-1}, X_0^{i-1}, B_{i-1}) = H(Y_i | B_{i-1}) = h(\tilde{P}_S(b+1)). \tag{11}$$

It follows from (9), (10), and (11) that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(Y_i|Y_0^{i-1},X_0^{i-1})$$
$$=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{b=0}^{n}h(\tilde{P}_S\,(b+1))P_{B_{i-1}}(b)$$
$$=\sum_{b=0}^{\infty}h(\tilde{P}_S\,(b+1))\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}P_{B_{i-1}}(b).$$

The claim follows since $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}P_{B_{i-1}}(b)=\pi_b$. ■

A. Memoryless Inter-arrivals

For memoryless arrivals, the Markov chain (A_i, B_i) described above will reduce to a Markov chain B_i with stationary distribution π_b. In order to apply Theorem 1 for general service distribution, we need the following lemma.

Lemma 2. *The M/G/1 bufferless SSTQ where λ is the rate of incoming tweets has limiting state probabilities*

$$\pi_b=\left(\frac{\lambda}{1-\lambda}\right)P(S>b)\pi_{-1}, b\geq 0$$

where $\pi_{-1}=(1-\lambda)[1+\lambda\,(E(S)-1)]^{-1}$.

Proof: Defining $q_i=P[S=i+1|S>i]$ as the service completion rate in state i, $p_i=1-q_i=P[S>i+1|S>i]$, the stationary probabilities for this Markov chain are given by $\Pi^T=\Pi^T\mathbb{P}$ where $\Pi^T=\begin{bmatrix}\pi_{-1} & \pi_0 & \pi_1 & \cdots\end{bmatrix}$, and

$$\mathbb{P}=\begin{bmatrix} 1-\lambda & \lambda & 0 & 0 & \cdots \\ (1-\lambda)\,q_0 & \lambda q_0 & p_0 & 0 & \cdots \\ (1-\lambda)\,q_1 & \lambda q_1 & 0 & p_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ (1-\lambda)\,q_i & \lambda q_i & 0 & \cdots & p_i\cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

■

Theorem 2. *For a general service time distribution with iid elements and independent from the memoryless inter-arrival time sequence,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(Y_i|Y_0^{i-1},X_0^{i-1})=\frac{H(S)\lambda}{1+\lambda\,(E(S)-1)}. \quad (12)$$

Proof: Using the results of Theorem 1 and Lemma 2,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(Y_i|Y_0^{i-1},X_0^{i-1})$$
$$=\sum_{b=0}^{\infty}h(\tilde{P}_S\,(b+1))\left(\frac{\lambda}{1-\lambda}\right)P(S>b)\pi_{-1}.$$

Based on Lemma 1,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(Y_i|Y_0^{i-1},X_0^{i-1})=\left(\frac{\lambda}{1-\lambda}\right)H(S)\pi_{-1}, \quad (13)$$

which, combined with Lemma 2, completes the proof. ■

In order to find the output entropy of the M/G/1 bufferless SSTQ, we need to define a Markov chain with states $M_i\in\{0,1,\cdots,n\}$ that specifies the number of units of time since the last departure. The limiting state distribution for this Markov chain can be shown to be

$$\gamma_m=P(D>m)/E(D). \quad (14)$$

Theorem 3. *The entropy rate of the sink node tweet sequence under memoryless tweet arrivals from the sender node is*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(Y_i|Y_1^{i-1})=\frac{H(D)}{E(D)}.$$

Proof: Since M_{i-1} is a deterministic function of Y_0^{i-1},

$$H(Y_i|Y_0^{i-1})=H(Y_i|Y_0^{i-1},M_{i-1}), \quad (15)$$

such that $H(Y_i|Y_0^{i-1}=y_0^{i-1},M_{i-1}=m)=h(\tilde{P}_D\,(m+1))$ for $m\geq 0$. As a result,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(Y_i|Y_0^{i-1})$$
$$=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{m=0}^{n-1}h(\tilde{P}_D\,(m+1))P_{M_{i-1}}(m)$$
$$=\sum_{m=0}^{\infty}h(\tilde{P}_D\,(m+1))\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}P_{M_{i-1}}(m). \quad (16)$$

It follows from (14) and the definition of limiting state probability that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}H(Y_i|Y_0^{i-1})=\sum_{m=0}^{\infty}h(\tilde{P}_D\,(m+1))\gamma_m$$
$$=\sum_{m=0}^{\infty}h(\tilde{P}_D\,(m+1))\frac{P(D>m)}{E(D)}.$$

The claim follows from Lemma 1. ■

The interpretation of Theorem 3 is as follows: If the length of Y sequence is n, the total entropy will be n times the entropy of one element in Y on average. Each symbol in the Y sequence will reveal part of the uncertainty in the inter-departure time D. With each departure, the length of corresponding inter-departure time is resolved. Since the time between two consecutive departures is on average $\mathrm{E}(D)$, entropy $H(D)$ is obtained every $\mathrm{E}(D)$ units of time using the concept of renewal reward theory which states that the time average of the amount of entropy you get is equal to the entropy per period divided by the expected value of the period.

Based on (12) and Theorem 3, the directed information rate from the source node A to sink node B is

$$I_\infty\,(DX\to Y)=\frac{H(D)}{\mathrm{E}(D)}-\frac{H(S)\lambda}{1+\lambda\,(\mathrm{E}(S)-1)}.$$

Moreover, the reverse directed information rate can be shown to be zero since the arrival process is independent of service process in the system.

For the M/G/1 queue, the idling times W_i are shifted geometric (λ) random variables $P_W(w)=\lambda\,(1-\lambda)^w\,, w=$

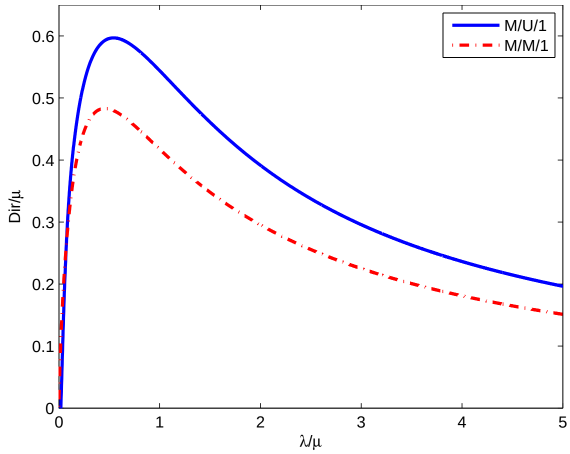

Fig. 2. The directed information rate $I_\infty(X \to Y)/\mu$ for the M/M/1 and M/U/1 (M/G/1 with uniform service times) queues. For each service discipline, we have observe that the normalized information rates are indistinguishable for $\mu \in \{0.01, 0.005, 0.001\}$.

$0, 1, 2, \cdots$ independent of S. In the special case of M/M/1, the service time is geometric with success probability μ so $P_S(s) = \mu(1-\mu)^{s-1}, s = 1, 2, 3, \cdots$. In this case, S has entropy $H(S) = h(\mu)/\mu$ and $D = W + S$ has PMF

$$P_D(d) = \frac{\mu\lambda}{\mu - \lambda}\left((1-\lambda)^d - (1-\mu)^d\right), \quad d = 1, 2, \cdots.$$

Another case to study is M/G/1 where S has uniform PMF $P_S(s) = 1/n$ for $s = 1, 2, \ldots, n$. Then, with $n = (2/\mu) - 1$, the PMF of $D = S + W$ is

$$P_D(d) = \begin{cases} (1-(1-\lambda)^d)/n & 1 \le d \le n, \\ (1-(1-\lambda)^n)(1-\lambda)^{d-n}/n & d > n. \end{cases}$$

Figure 2 illustrates the directed information rate $I_\infty(X \to Y)/\mu$ for the case that the time slot is 1 second and the time scale represents tweeting with $10^{-4} \le \lambda \le 10^{-2}$ which corresponds to approximately 8 tweets per day to 860 tweets per day. Moreover, average thinking time is $1/\mu$ where $\mu \in \{0.01, 0.005, 0.001\}$. In the figure, we observe that the normalized directed information rate as a function of the normalized arrival rate λ/μ is insensitive to the choice of μ; however, analytic verification of this observation remains to be completed. Furthermore, it can be seen that when λ is close to 0, the directed information rate is negligible since the source node rarely tweets so it does not have high influence. By increasing λ, the source node is tweeting more, so the measured influence will increase. However, when the sender tweets much faster than the receiver node can process, many tweets are deleted, ultimately leading to a decrease in the directed information rate. Another important point is that for the same tweet arrival rate λ, increasing μ will increase the information rate. Intuitively, increasing μ could be interpreted as an increase in a person's willingness to read the tweets faster. The apparent similarity of the figures suggests that timing information in this setting appears to be relatively insensitive to the service time distributions.

IV. Conclusion

In this paper, we have investigated the role of timing information in understanding influence in social networks. For a simple two-node Twitter network in which one nodes influences the other, we have related the problem of inference of influences to that of packet transmissions in bufferless single-server timing queues. We have computed the directed information rate in this case and interpreted the effects of changes in input/output timings on the directed information rate. Future work in this direction involves understanding the role of memory in the timing information of the influencing user's tweets, extensions to the case when both users exert some influence on each other, and generalizations of this problem to larger social networks.

References

[1] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *Journal of computational neuroscience*, vol. 30, no. 1, pp. 17–44, 2011.

[2] C. Quinn, N. Kiyavash, and T. Coleman, "Directed information graphs," *Arxiv preprint arXiv:1204.2003*, 2012.

[3] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 5, no. 4, p. 21, 2012.

[4] G. Ver Steeg and A. Galstyan, "Information transfer in social media," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 509–518.

[5] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," in *Machine learning and knowledge discovery in databases*. Springer, 2011, pp. 18–33.

[6] N. Kayastha, D. Niyato, P. Wang, and E. Hossain, "Applications, architectures, and protocol design issues for mobile social networks: A survey," *Proceedings of the IEEE*, vol. 99, no. 12, pp. 2130–2158, Dec 2011.

[7] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.

[8] S. A. Myers and J. Leskovec, "On the convexity of latent social network inference," *threshold*, vol. 9, p. 20, 2010.

[9] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[10] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology (ETH), Switzerland, 1998.

[11] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, 1990, pp. 303–305.

[12] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

[13] G. K. Gomez-Rodriguez, M. and B. Schoelkopf, "Quantifying information overload in social media and its impact on social contagions," in *Eighth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2014.

[14] M. Tavan, R. D. Yates, and W. U. Bajwa, "Bits through bufferless queues," in *51st Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2013.

[15] ——, "Capacity analysis of a discrete-time bufferless timing channel," in *48th Annual Conference on Information Sciences and Systems*. IEEE, 2014.

[16] V. Anantharam and S. Verdu, "Bits through queues," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 4–18, 1996.

[17] A. S. Bedekar and M. Azizoglu, "The information-theoretic capacity of discrete-time queues," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 446–461, 1998.