An information-theoretic approach to study activity driven networks

Christian Bongiorno
Dipartimento di Automatica
e Informatica
Politecnico di Torino
Torino, Italy
Email: christian.bongiorno@polito.it

Alessandro Rizzo
Dipartimento di Elettronica
e Telecomunicazioni
Politecnico di Torino
Torino, Italy
Email: alessandro.rizzo@polito.it

Maurizio Porfiri
Department of Mechanical
and Aerospace Engineering
New York University
Tandon School of Engineering
Brooklyn NY, USA
Email: mporfiri@nyu.edu

***Abstract*—In this work, we leverage the information-theoretic notion of transfer entropy theory to study causal information flow in epidemic spreading over temporal networks. An improved understanding of causal information flow may lead to the early detection of population segments that should be monitored or immunized to enhance epidemic containment. We focus on activity driven networks, which constitute a powerful and elegant paradigm to capture the inherent time-varying nature of contacts and population heterogeneity. Our preliminary results confirm our intuition that individuals who have a higher propensity in contacting others are responsible for the largest information transfer. Moreover, we find that epidemic parameters such as the probability of infection and recovery may dominate the spreading phenomenon over heterogeneities in the contact formation.**

I. Introduction

Temporal networks have recently gained traction in the research community to offer a more realistic description of collective dynamics in engineering and science [1], [2]. A temporal network is a network where the links are continuously created and removed as time progresses. One of the classical approaches to study temporal networks is to integrate them over time to form a static network, encapsulating all the links that have at least appeared once during the time evolution. Then, the integrated network is studied using classical tools in network theory [3], [4]. The recent availability of large time-resolved datasets, collected over long time spans, allows for unprecedented empirical analysis and the development of theoretic frameworks for temporal networks [5].

Activity driven networks (ADNs) are a promising modeling paradigm in temporal networks, which naturally accounts for the heterogeneity of nodes in the propensity to create contacts with their peers. ADNs provide a simple and elegant means to model the presence of *hubs* that, in the temporal case, are nodes that are most likely to generate contacts with the rest of the network [6]. Each node of an ADN is characterized by a constant activity rate, which quantifies the probability per unit time that a node is *active,* namely, capable of generating links.

Major contributions in ADN-related studies have been directed toward understanding the dynamics of epidemic diffusion processes. A seminal work by Perra *et al.* [6] introduced the notion of ADNs and explored the Susceptible-Infected-Susceptible (SIS) model. Issues on the control of the contagion and the critical immunization threshold have been tackled in [7]. The influence of individual behavior on the epidemic outbreak has been elucidated by our group in [8]. Further efforts have been devoted to map the Susceptible-Infected-Removed (SIR) model to percolation problems [9]. The effect of node memory has been treated in [10], and community structures have been investigated in [11], [12]. Most of these studies are based on extensive simulation campaigns: an alternative formulation of ADNs based on a continuous-time model and a discrete distribution of activity rate has been recently established by our group in [13], [14]. This approach favors a complete analytical treatment and the prediction of the entire dynamics of the epidemic spreading.

Much effort is currently being devoted to steer ADN models from the simulation of simple epidemic processes to comprehensive characterization of realistic spreading. For example, in [15], [16], we have proposed an ADN model with behavioral patterns and the adoption of time-varying containment protocols to simulate and predict the dynamics of the 2014-2015 Ebola Virus Disease in Liberia. The proposed model, calibrated on available field data, can predict the spreading dynamics over a long time horizon and supports what-if analysis on the effect of the introduction of timely containment interventions.

Toward implementing effective containment strategies, epidemic models would benefit from analysis tools to detect individuals that are most responsible for the epidemic propagation. This issue has been widely studied in the literature for static networks, using two main approaches: preventive and reactive immunization. Preventive immunization leverages information about individuals obtained before the epidemic inception [17], [18]. Within preventive approaches, targeted immunization uses information on topological properties of individuals (e.g., node degree and betweenness centrality). On the other hand, reactive immunization focuses on the dynamics of the epidemic spreading, accounting for dynamical aspects of the network and of the epidemic itself to detect the most important individuals to be vaccinated [19], [20].

Further work has been devoted to detect social network sensors, a restricted subset of individuals to be monitored to

get ahead of the epidemics, before it hits a wide portion of the population [21], [22].

Similar studies on temporal networks are being recently conducted, leveraging temporal correlation techniques [23], or the relationship between vaccination strategies and an optimal choice of sentinels toward an early detection of the outbreak [24]. Although promising results are being obtained, research in this field is still in its infancy.

Here, we leverage an information-theoretic approach toward a preliminary study of the causal information flow in ADNs. Revealing patterns in causal information flow may lead to the introduction of new, data-driven metrics to select key individuals in an epidemic process. In an information-theoretic sense, we seek to elucidate how uncertainty propagates in the system and attempt to learn the causal basis of the propagation [25], [26]. Thus, a large information flow from a node to others in the network may tell us that such a specific node is critical for the outbreak and should be promptly monitored or immunized.

In this work, we use transfer entropy (TE) to study causal information flow between class of nodes [27]. Transfer entropy has been used to tackle a wide range of applications, such as in finance [28], neuroscience [29], biology [30], and public policies [31]. In our application, information flow quantified by TE corresponds to the reduction of the uncertainty in the prediction of the future in the epidemic dynamics of a class of individuals due to the knowledge of the past epidemic dynamics of another class of individuals.

This paper reports the results of a preliminary study on the topic of information transfer in ADNs, aiming at quantifying information flow during an epidemic process and its dependence on *i)* the propensity of nodes to make contact with others; and *ii)* the epidemic parameters. The paper is organized as follows: Section II provides the necessary background on ADNs, epidemic processes on ADNs, and transfer entropy. Simulation results are reported and commented in Section III, and our conclusions are summarized in Section IV

II. Problem Statement

A. Activity-Driven Networks

An ADN is a temporal network, whose dynamics is typically described in discrete time. The n nodes that comprise the network are characterized by an individual activity rate. Such rates are collected in a vector $a = [a_1, a_2, \cdots, a_n]$, where $a_i \in [0, 1]$ is the probability that node i becomes *active* in a time-step. Activation rates are drawn by a given probability distribution. It has been observed that a power law distribution with exponent in the range $[2, 3]$ offers a satisfactory fit for real socio-technical systems [6], [16].

When a node is active, it connects randomly to m other nodes in the network. An epidemic process is then run on the resulting, possibly disconnected, network. Links formed in a time-step are removed and the process resumes to the new time-step.

Network nodes do not have memory of their past history, such that the link formation at each time-step is independent from past instances of the network topology. Thus, the integration of the network along T time-steps may produce a degree distribution with a broad variability, comprising hubs that correspond to nodes with the highest activation rates [6].

Figure 1 illustrates the temporal integration process of an ADN with $n = 15$ nodes and $m = 3$ links per active node. We observe that hubs tend to form already from $t = 4$ in correspondence of nodes with the highest acvitity, marked in red in the figure.

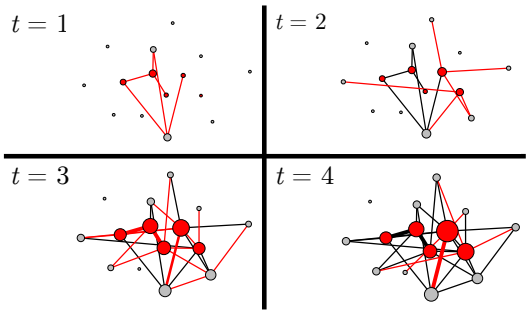

Fig. 1. Temporal integration of the first 4 time-steps of the link formation process of an ADN with $n = 15$ nodes and $m = 3$ link per active node. Red links indicate the connections that are created in the current time-step. Black links result from the integration over the previous time-steps. Red nodes have high activity rate $a_h = 0.3$ and gray nodes have low activity rate $a_l = 0.05$. Node size is proportional to the node degree.

B. Susceptible-Infected-Susceptible Model in an ADN

In a SIS model, each node in a time-step t is characterized by two possible states, susceptible or infected. At time-step t, once the network of contacts is formed according to the ADN process, a susceptible node i can contract the disease from a neighbor j, which was in the infected state at time $t - 1$, with probability λ. Moreover, a node that was in the infected state at time $t - 1$ can recover at time t with probability μ. An example of an SIS process over an ADN is illustrated in Fig. 2.

A critical *epidemic threshold* for the ratio λ/μ exists, such that the epidemic becomes endemic when the threshold is passed. As a result, after a transient phase, the average number of infected nodes fluctuates around a stationary value, which is also known as the *endemic state*. For a SIS process on ADNs, such a critical threshold is regulated by the following expression [6], [8], [13], [14]:

$$\frac{\lambda}{\mu} > \frac{1}{m} \frac{1}{\langle a \rangle + \sqrt{\langle a^2 \rangle}}, \tag{1}$$

where $\langle \cdot \rangle$ indicates average, and a^2 is the activity vector a squared element-wise. The expected increment of infected nodes with a given activity rate z in a time-step of unitary length is described by the following equation [6]:

$$I_z^{t+1} = (1 - \mu) I_z^t + \lambda m (N_z - I_z^t) \left(z \sum_{z'} \frac{I_{z'}^t}{N} + \sum_{z'} \frac{z' I_{z'}^t}{N} \right), \tag{2}$$

where N_z is the number of nodes with activity rate z, and z' is the set of nodes with activity rate different from z. We observe

that the number of infected nodes in Eq. (2) is a nonlinear combination of the number of infected nodes with different activity rates at the previous time-step. This consideration makes TE suitable for our purposes, whereas other data-driven metrics, like Granger causality [25], may not be used to tackle these nonlinear phenomena.

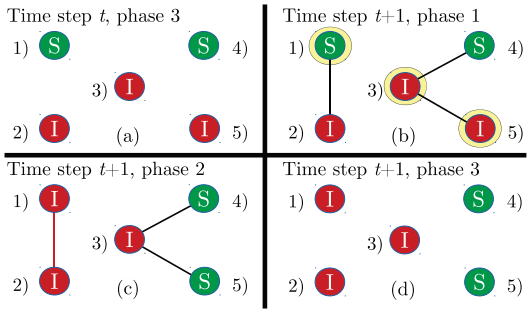

Fig. 2. An SIS epidemic model evolving on an ADN with N = 5 nodes and m = 1 links per active node. Nodes' health states are encircled, and active nodes are highlighted by a thick yellow border. (a) At the last phase of time t, the ADN is disconnected and nodes 2, 3, and 5 are infected. In the first phase of time-step $t+1$ (b) nodes 1, 3, and 5 become active and contact nodes 2, 4, and 3, respectively; (c) the epidemic process evolves, so that node 2 infects node 1, node 3 remains in the infected state, node 4 remains in the susceptible state, and node 5 recovers; and (d) time t has elapsed and all the network edges are removed before a new time increment is initiated.

C. Transfer Entropy Between Activity Classes

We define an activity class x as the subset of the network nodes whose activity lays within a fixed real range $[w_x^-, w_x^+]$. We extend the notation in Eq. (2) by defining the aggregate variables I_x^t and I_y^t, which represent the number of infected nodes at time-step t that belong to class x and y respectively. For a generic class, indicated with "$\bullet$", the sample space of $I_\bullet^t$ is indicated with $\mathcal{I}_\bullet$. Such a sample space is an integer interval $0, \dots N_\bullet$, where $N_\bullet$ is the number of nodes in the activity class.

Transfer entropy from class x to class y quantifies the reduction of the uncertainty in the prediction of the future of the dynamics of class y due to the knowledge of the dynamics of class x, and is defined through three steps [27]. First, conditioned entropy within class y measures the uncertainty in I_y^{t+1} given the knowledge of I_y^t. It is defined as

$$H(I_y^{t+1}|I_y^t) = \\ = \sum_{\alpha\in\mathcal{I}_y,\,\beta\in\mathcal{I}_y} p(I_y^{t+1}=\alpha, I_y^t=\beta)\log\frac{p(I_y^t=\beta)}{p(I_y^{t+1}=\alpha, I_y^t=\beta)}. \tag{3}$$

Then, the entropy of class y conditioned to both I_x^t and I_y^t is

$$H(I_y^{t+1}|I_y^t, I_x^t) = \\ = \sum_{\alpha\in\mathcal{I}_y,\,\beta\in\mathcal{I}_y,\gamma\in\mathcal{I}_x} p(I_y^{t+1}=\alpha, I_y^t=\beta, I_x^t=\gamma)\times \\ \times\log\frac{p(I_y^t=\beta, I_x^t=\gamma)}{p(I_y^{t+1}=\alpha, I_y^t=\beta, I_x^t=\gamma)}. \tag{4}$$

Finally, TE from class x to class y is

$$TE^{x\to y} = H(I_y^{t+1}|I_y^t) - H(I_y^{t+1}|I_y^t, I_x^t) = \\ = \sum_{\alpha\in\mathcal{I}_y,\,\beta\in\mathcal{I}_y,\gamma\in\mathcal{I}_x} p(I_y^{t+1}=\alpha, I_y^t=\beta, I_x^t=\gamma)\times \\ \times\log\frac{p(I_y^{t+1}=\alpha|I_y^t=\beta, I_x^t=\gamma)}{p(I_y^{t+1}=\alpha|I_y^t=\beta)}. \tag{5}$$

The following section demonstrates how TE can be used to gain insight in the epidemic spreading.

III. Results

We consider an ADN composed of $N = 900$ nodes divided in three distinct classes of activity, each characterized by a singleton value: $N_h = 300$ nodes with high activity $a_h = 0.9$; $N_m = 300$ nodes with moderate activity $a_m = 0.6$; and $N_l = 300$ nodes with low activity $a_l = 0.3$. We set the number of links generated by an active node to $m = 1$.

We execute the first set of simulations by fixing the SIS infection and recovery probabilities to $\lambda = 0.8$ and $\mu = 0.4$, respectively. A simple application of Eq. (1) reveals that these values are above the epidemic threshold. A seed of 10% of randomly distributed nodes is initially placed. We perform $1,000$ independent simulations of the SIS process, with simulation length equal to $T = 100,000$ time-steps. In this work, we focus on the stationary state of the epidemic spreading. To this aim, the first $1,000$ time samples are removed from each time series.

We estimate the TE between each pair of classes. Since numerical issues may arise when the cardinality of the sample spaces is large, we aggregate the sample spaces into smaller ones. Specifically, we collapse the $N_\bullet = 300$ possible states of $\mathcal{I}_\bullet$ into $N'_\bullet = 30$ aggregated states, by binning $\mathcal{I}_\bullet$ into groups of 10 elements. This procedure has the drawback of reducing the amount of information that we are able to extract from data, however, it is necessary to reduce the error in the estimation of the TE, with time series of a length of $T = 100,000$ steps.

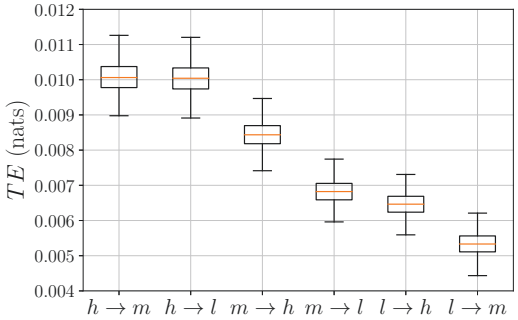

Fig. 3. TE among each pair of activity classes estimated on $1,000$ independent simulations. The symbols l, m, h correspond to low, middle, and high activity, respectively. Boxes represent the 25-75 percentile range, whiskers represent the extreme values, and orange lines the median one.

Figure 3 illustrates a box-plot of TE between each pair of classes. In agreement with our intuition, we observe that the

highest activity class elicits the largest information transfer, and that information transfer decreases with activity. This suggests that nodes with high activity are more influential and knowledge about their dynamics is critical to reconstruct the entire spreading process from partial knowledge. We also note that the strength of the interaction between activity classes increases with the level of activity. In other words, from Fig. 3, we find that TE from medium to high activity classes is higher than TE from medium to low activity classes, and TE from low to high activity classes is higher than TE from low to medium activity classes. Finally, we note that despite activity classes are equally spaced, the corresponding TE values are not.

A second set of simulations is devoted to investigate how TE from a class to all the other classes is influenced by the SIS parameters of the infection, namely, the probability of infection λ and that of recovery μ. Toward this aim, we define the average TE from a class x to all the other classes as follows:

$$\langle TE^{x\to\bullet}\rangle = \frac{\sum_{y\neq x} TE^{x\to y}}{n_c - 1}, \tag{6}$$

where n_c is the number of classes. In this second set of simulations, we fix first $\mu = 0.4$ and we let λ vary in the interval $[0.42, 1.0]$. Then, we fix $\lambda = 0.8$ and we let μ vary in the interval $[0.10, 0.80]$. The range of the parameters guarantee the attainment of the endemic state, since the epidemic threshold in Eq. (1) is always passed.

Figure 4 illustrates the trend of $\langle TE^{x\to\bullet}\rangle$ from one class to all the others, as a function of the infection probability λ. Simulation results confirm that, even in this aggregated formulation, transfer entropy from higher activity classes is always larger than that from lower activity classes. Also, we observe that $\langle TE^{x\to\bullet}\rangle$ decreases with increasing values of λ. This is due to the fact that a high value of the contagion probability leads to a saturation to high values of I_x^t for all the classes, such that the information content is mostly driven by the infection probability rather than the node activity. This explanation is also confirmed by observing that the separation between the three entropy values for each class tends to reduce for higher values of λ, that is, the activity rate plays a secondary role in the propagation of the infection.

Figure 5 illustrates the trend $\langle TE^{x\to\bullet}\rangle$ from one class to all the others, as a function of the recovery probability μ. Opposite to the trend if Fig. 4, in Fig. 5 we find that $\langle TE^{x\to\bullet}\rangle$ increases for high values of μ. For low values of μ, the large average time needed for a node to recover from the disease leads to a condition with most of the nodes permanently infected. In such a condition, the activity value has a modest effect on the epidemic dynamics. On the other hand, when μ attains large values, the activity takes a key role in the information spreading and $\langle TE^{x\to\bullet}\rangle$ from higher activity classes is larger than that from lower activity classes.

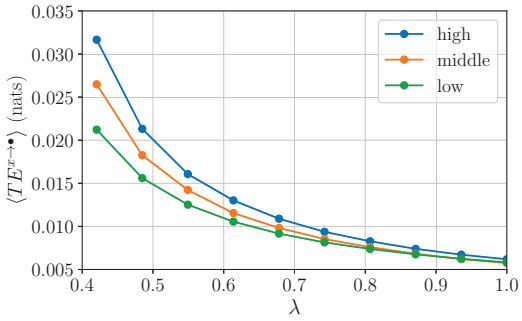

Fig. 4. Average TE from each class x to all the other classes as a function of the infection probability λ. We fix $\mu = 0.4$ and $\lambda \in [0.42, 1.0]$. Each point is an average of 100 independent simulations.

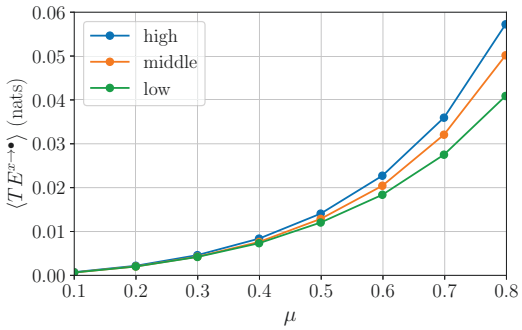

Fig. 5. Average TE from each class x to all the other classes as a function of the recovery probability μ. We fix $\lambda = 0.8$ and $\mu \in [0.10, 0.80]$. Each point is an average of 100 independent simulations.

IV. Conclusions

Here, we have established an information-theoretic approach to study causal information flow in epidemic spreading in temporal networks. We have proposed the use of transfer entropy to identify critical network nodes more influential in the spreading process. Activity-driven networks have been selected as a modeling paradigm for temporal networks, as they are able to encapsulate the inherent heterogeneity in the propensity to generate contacts with others.

Our simulations confirm intuition, in that network nodes that are most likely to generate contacts with others are the most influential in the epidemic spreading. On the other hand, when epidemic parameters such as the infection or the recovery probability attain extreme values, the individual propensity of a node to make contacts assumes a secondary role. Further work will deal with an analytical characterization of casual information flow, immunization strategies based on information transfer, and the assessment of the transient phase of the epidemic spreading.

Acknowledgment

This work was supported by the National Science Foundation under grant No. CMMI-1561134, the Army Research Office under grant No. W911NF-15-1-0267, with Drs. A. Garcia and S.C. Stanton as program managers, and by Compagnia di San Paolo.

REFERENCES

[1] P. Holme and J. Saramäki, "Temporal networks," *Physics Reports*, vol. 519, pp. 97–125, 2012.

[2] I. Belykh, M. Di Bernardo, J. Kurths, and M. Porfiri, "Evolving dynamical networks," *Physica D: Nonlinear Phenomena*, vol. 267, pp. 1–6, 2014.

[3] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7332–7336, 2007.

[4] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, "The complex network of global cargo ship movements," *Journal of the Royal Society Interface*, vol. 7, no. 48, pp. 1093–1103, 2010.

[5] J. Tang, C. Mascolo, M. Musolesi, and V. Latora, "Exploiting temporal complex network metrics in mobile malware containment," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a World of Wireless, Mobile, and Multimedia Networks*. IEEE, 2011, pp. 1–9.

[6] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, "Activity driven modeling of time varying networks." *Scientific Reports*, vol. 2, p. 469, 2012.

[7] S. Liu, N. Perra, M. Karsai, and A. Vespignani, "Controlling contagion processes in activity driven networks," *Physical Review Letters*, vol. 112, no. 11, p. 118702, 2014.

[8] A. Rizzo, M. Frasca, and M. Porfiri, "Effect of individual behavior on epidemic spreading in activity driven networks," *Physical Review E*, vol. 90, p. 042801, 2014.

[9] M. Starnini and R. Pastor-Satorras, "Temporal percolation in activity-driven networks," *Physical Review E*, vol. 89, p. 032807, 2014.

[10] K. Sun, A. Baronchelli, and N. Perra, "Contrasting effects of strong ties on sir and sis processes in temporal networks," *European Physical Journal B*, vol. 88, pp. 1–8, 2015.

[11] M. Liu, W. Wang, Y. Liu, M. Tang, S. Cai, and H. Zhang, "Social contagions on time-varying community networks," *Physical Review E*, vol. 95, no. 5, p. 052306, 2017.

[12] M. Nadini, K. Sun, E. Ubaldi, M. Starnini, A. Rizzo, and N. Perra, "Epidemic spreading in modular time-varying networks," *arXiv preprint arXiv:1710.01355*, 2017.

[13] L. Zino, A. Rizzo, and M. Porfiri, "Continuous-time discrete-distribution theory for activity-driven networks," *Physical Review Letters*, vol. 117, p. 228302, 2016.

[14] ——, "An analytical framework for the study of epidemic models on activity driven networks," *Journal of Complex Networks*, p. cnx056, to appear, 2017.

[15] A. Rizzo and M. Porfiri, "Toward a realistic modeling of epidemic spreading with activity driven networks," in *Temporal Network Epidemiology*. Springer, 2017, pp. 317–342.

[16] A. Rizzo, B. Pedalino, and M. Porfiri, "A network model for Ebola spreading," *Journal of Theoretical Biology*, vol. 394, pp. 212 – 222, 2016.

[17] R. Pastor-Satorras and A. Vespignani, "Immunization of complex networks," *Physical Review E*, vol. 65, no. 3, p. 036104, 2002.

[18] L. Hébert-Dufresne, A. Allard, J. Young, and L. Dubé, "Global efficiency of local immunization on complex networks," *Scientific Reports*, vol. 3, 2013.

[19] F. Altarelli, A. Braunstein, L. Dall'Asta, J. Wakeling, and R. Zecchina, "Containing epidemic outbreaks by message-passing techniques," *Physical Review X*, vol. 4, no. 2, p. 021024, 2014.

[20] E. Alfinito, M. Beccaria, A. Fachechi, and G. Macorini, "Reactive immunization on complex networks," *Europhysics Letters*, vol. 117, no. 1, p. 18002, 2017.

[21] N. Christakis and J. Fowler, "Social network sensors for early detection of contagious outbreaks," *PloS one*, vol. 5, no. 9, p. e12948, 2010.

[22] H. Shao, K. Hossain, H. Wu, M. Khan, A. Vullikanti, B. Prakash, M. Marathe, and N. Ramakrishnan, "Forecasting the flu: designing social network sensors for epidemics," *arXiv preprint arXiv:1602.06866*, 2016.

[23] S. Lee, L. Rocha, F. Liljeros, and P. Holme, "Exploiting temporal network structures of human interaction to effectively immunize populations," *PloS one*, vol. 7, no. 5, p. e36439, 2012.

[24] Y. Bai, B. Yang, L. Lin, J. Herrera, Z. Du, and P. Holme, "Optimizing sentinel surveillance in temporal network epidemiology," *Scientific Reports*, vol. 7, no. 1, p. 4804, 2017.

[25] L. Barnett, A. Barrett, and A. Seth, "Granger causality and transfer entropy are equivalent for Gaussian variables," *Physical Review Letters*, vol. 103, no. 23, p. 238701, 2009.

[26] P. Duan, F. Yang, T. Chen, and S. L. Shah, "Direct causality detection via the transfer entropy approach," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 6, pp. 2052–2066, 2013.

[27] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, p. 461, 2000.

[28] O. Kwon and J. Yang, "Information flow between stock indices," *Europhysics Letters*, vol. 82, no. 6, p. 68003, 2008.

[29] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy - a model-free measure of effective connectivity for the neurosciences," *Journal of computational neuroscience*, vol. 30, no. 1, pp. 45–67, 2011.

[30] S. Butail, F. Ladu, D. Spinello, and M. Porfiri, "Information flow in animal-robot interactions," *Entropy*, vol. 16, no. 3, pp. 1315–1330, 2014.

[31] M. Porfiri and M. Marin, "Information flow in a model of policy diffusion: an analytical study," *IEEE Transactions on Network Science and Engineering*, 2017.