Tracking Infection Diffusion in Social Networks: Filtering Algorithms and Threshold Bounds

Vikram Krishnamurthy*, *Fellow, IEEE*   Sujay Bhatt†   Tavis Pedersen‡

Abstract—This paper deals with the statistical signal processing over graphs for tracking infection diffusion in social networks. Infection (or Information) diffusion is modeled using the Susceptible-Infected-Susceptible (SIS) model. Mean field approximation is employed to approximate the discrete valued infected degree distribution evolution by a deterministic ordinary differential equation for obtaining a generative model for the infection diffusion. The infected degree distribution is shown to follow polynomial dynamics and is estimated using an exact non-linear Bayesian filter. We compute posterior Cramér-Rao bounds to obtain the fundamental limits of the filter which depend on the structure of the network. Considering the time-varying nature of the real world networks, the relationship between the diffusion thresholds and the degree distribution is investigated using generative models for real world networks. In addition, we validate the efficacy of our method with the diffusion data from a real-world online social system, Twitter. We find that SIS model is a good fit for the information diffusion and the non-linear filter effectively tracks the information diffusion.

Index Terms—Social Networks, Cramér-Rao bounds, mean-field dynamics, non-linear Bayesian filter, Twitter dataset, diffusion threshold, stochastic dominance.

I. INTRODUCTION

Statistical signal processing on graphs is an emerging field in which the structural properties of the graph are utilized to derive statistical inference algorithms. As described in [1], there is a wide range of social phenomena such as diffusion of technological innovations, cultural fads, and economic conventions [2] where individual decisions are influenced by the decisions of others. In this paper, we consider social networks represented as graphs and we are interested in analyzing the manner in which information (or infection) spreads through the network. A large body of research on social networks has been devoted to the diffusion of information (e.g., ideas, behaviors, trends) [3], and particularly on finding a set of target nodes so as to maximize the spread of a given product [4], [5].

Organization and Main Results

Sec. II presents the *Susceptible Infected Susceptible* (SIS) model for infection diffusion in the network. The key result is that the mean field dynamics approximates the discrete-valued infected degree distribution evolution by a deterministic ordinary differential equation. The mean field dynamics yield

a tractable model for Bayesian filtering in order to estimate the infected degree distribution given a sampling procedure for the social network. Although, the formulation and mean field dynamics are known, the proof presented has tutorial value and uses a martingale-based Azuma-Hoeffding inequality. From a signal processing point of view, the mean field dynamics has an interesting interpretation: it resembles a stochastic approximation algorithm; however, in our case, it constitutes an underlying generative model instead of an algorithm.

The mean field dynamics of Sec. II yields a dynamical system whose state (infected degree distribution of network) evolves with polynomial dynamics. Sec. III uses a recent result in Bayesian filtering to obtain an exact (finite dimensional filter) for the the infected degree distribution given noisy observations. We examine via numerical examples and posterior Cramér-Rao lower bounds, how state estimation over large networks is affected by the network. Numerical examples we demonstrate the difference in performance between power law (Scale Free) and Erdős-Rényi graphs.

The classical SIS model assumes a fixed underlying social network. Sec. IV analyzes the diffusion threshold of a SIS model when the social network evolves over time. Since information diffusion occurs at a faster time scale compared to forming connections in social networks, we consider a two time scale formulation: the degree distribution of the underlying network changes on a slow time scale and the infection diffuses over a faster time scale. Our results generalize the results in [1], where the underlying network was assumed to be fixed. By using results in stochastic dominance we show that in a preferential attachment model for a randomly evolving graph, the infection diffusion threshold decreases with the attachment probability. To the best of our knowledge, this is new.

Sec. V illustrates the SIS model and the performance of the Bayesian filter on simulated data and examines the sensitivity of the filter to the underlying graph model (Erdős-Rényi vs Scale Free). Sec. V also presents a detailed example using a real Twitter dataset. It is shown via a goodness of fit that SIS is a reasonable model for information propagation in the Twitter dataset and that the infected degree distribution can be tracked satisfactorily over time via the Bayesian filter.

Literature

Susceptible-Infected-Susceptible (SIS) models for diffusion of information in social networks have been extensively studied in [1], [6], [7], [8], [9], [10], [11], [12] to model, for example, the adoption of a new technology in a consumer

---

* and † are with the Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850. ‡ is with the Department of Electrical and Computer Engineering at the University of British Columbia, Vancouver, Canada, V6T1Z4.

Email: *vikramk@cornell.edu, †sh2376@cornell.edu and ‡tpedersen@ece.ubc.ca

market. The literature in SIS and related models for infection diffusion is vast. Our contribution in this paper is focussed on estimating (tracking) the evolving infected distribution using the mean field dynamics (MFD) as a generative model. The mean field dynamics yield a state space model with polynomial dynamics and sampling the network yields noisy measurements of the infected degree distribution.

In this paper, we assume that measurements of the infected degree distribution are obtained by sampling the network. There are various network sampling methodologies studied in the literature, see [12], [13], [14], [15], [16], [17], [18], [19], [20]. We consider two popular sampling methods: Uniform sampling and Respondent Driven Sampling (RDS) [18], [19]. It is shown that under reasonable conditions, by the central limit theorem, that the estimate of the probability that a node is infected in a large population (given its degree) is Gaussian under these sampling methods. Filtering algorithms for polynomial dynamical systems in Gaussian noise were recently developed in [21]. It was shown in [21] that one can devise a finite dimensional filter (based on the Kalman filter) to compute the conditional mean estimate. Therefore, for MFD, it is possible to obtain an exact Bayesian filtering algorithm for tracking the infected degree distribution.

To determine a lower bound for the mean square error of the optimal filter, we compute the posterior Cramér-Rao bounds [22]. It is observed that the performance of the optimal filter is relatively insensitive to the underlying network degree distribution.

The main result in Sec. IV deals with analyzing how the SIS model is affected when the underlying social network evolves with time. On networks having fixed degree distribution, [1] identified conditions under which a network is susceptible to an epidemic using a mean-field approach and provided a closed form solution for the infection diffusion threshold. The diffusion properties of networks was investigated using stochastic dominance of their underlying degree distributions like in [23], [24]. However, as real world networks are time evolving, we extend the analysis of diffusion thresholds to time evolving networks using generative models for the underlying network evolution. [25] provides a stochastic approximation algorithm and analysis on a Hilbert space for tracking the degree distribution of evolving random networks with a duplication-deletion model. There are various generative models for large real world networks in the literature, see [26], [27], [25], and the references therein. In this paper, we consider one such model, namely, the preferential attachment model [26], and analyze the connection between the diffusion threshold and the parameters that dictate evolution. The primary motivation for choosing a preferential attachment graph is that it is the simplest graph whose steady state distribution obeys a power law [26], that commonly arises in most real world networks, see [28], [29], [30].

Finally, in Sec. V, we construct and evaluate the SIS model on a Twitter dataset to track the diffusion of information over time using the filtering algorithms developed in Sec III. SIS model and its application to the diffusion of information on Twitter has been studied in [31], [32]. While these papers analyze the effectiveness of SIS on Twitter and other social media, in this paper we not only show via goodness of fit that SIS is a reasonable model, but also show that a filter can satisfactorily track the infection diffusion modeled using mean field dynamics.

II. SIS Model and Mean Field Dynamics

Consider a social network where the states of individual nodes evolve over time as a probabilistic function of the states of their neighbors. This evolution can be seen as distributed information processing by the individual nodes to estimate an underlying state process. We model this evolution as infection diffusion over a fast time scale using a SIS model, where the information about the underlying state is accessed using network sampling procedures described in Sec. II-B.

A. SIS Model and Mean Field Dynamics

This section discusses the population model and mean field dynamics for the diffusion of information in a social network. The formulation closely follows [1]. Consider a social network represented by graph G:

$$G = (V, E), \text{ where } V = \{1, \ldots, M\}, \text{ and } E \subseteq V \times V. \quad (1)$$

Here, V denotes the set of M vertices (users) and E denotes the set of edges (relationships). The degree of a node m is its number of neighbors:

$$D^{(m)} = |\{n \in V : m, n \in E\}|,$$

where $|\cdot|$ denotes cardinality.

Let $M(l)$ denote the number of nodes in the social network G with degree l, and let the degree distribution $\rho(l)$ specify the fraction of nodes with degree l. That is, for $l = 0, 1, \ldots, L$,

$$M(l) = \sum_{m \in V} I(D^{(m)} = l), \quad \rho(l) = \frac{M(l)}{M}.$$

Here, $I(\cdot)$ denotes the indicator function. Since $\sum_l \rho(l) = 1$, $\rho(l)$ can be viewed as the probability that a node selected randomly on V has connectivity l.

Assume each node m has two possible states,

$$s_n^{(m)} \in \{1 \text{ (infected) }, 2 \text{ (susceptible) }\}.$$

Let $x_n(l)$ denote the fraction of nodes with degree l at time n that are infected. We call the L dimensional vector x_n as the *infected population state* at time n. So

$$x_n(l) = \frac{1}{M(l)} \sum_m I(D^{(m)} = l, s_n^{(m)} = 1), \quad l = 1, \ldots, L. \quad (2)$$

1) SIS Model: The SIS model assumes that the infected population evolves as follows: If node m has degree $D^{(m)} = l$, then it jumps from state i to j with probabilities

$$P_{ij}(l, a) = \mathbb{P}\left(s_{n+1}^{(m)} = j | s_n^{(m)} = i, D^{(m)} = l, F_n^{(m)} = a\right)$$
$$i, j \in \{1, 2\}. \quad (3)$$

Here, $F_n^{(m)}$ denotes the number of infected neighbors of node m at time n. In words, SIS model assumes that the transition probability of a node depends on its degree and the number

of infected neighbors.

The infected population state is updated as

$$x_{n+1}(l) = x_n(l) + \frac{1}{M(l)}[I(s_{n+1}^{(l)} = 1, s_n^{(l)} = 2)$$
$$- I(s_{n+1}^{(l)} = 2, s_n^{(l)} = 1)] \quad (4)$$

According to (4), the infected population state changes by $\frac{1}{M(l)}$ at every timestep depending on the transition probabilities.

The following statistic forms a convenient parametrization of the transition probabilities of x_n. Define $\alpha(x_n)$ as the probability that a uniformly sampled link in the network at time n has at least one node that is infected. We call $\alpha(x_n)$ as the *infected link probability*. Clearly,

$$\alpha(x_n) = \frac{\sum_{l=1}^{L}(\text{\# of links from infected node of degree } l)}{\sum_{l=1}^{L}(\text{\# of links of degree } l)}$$
$$(5)$$
$$= \frac{\sum_{l=1}^{L} l\,\rho(l)\,x_n(l)}{\sum_{l}^{L} l\,\rho(l)}.$$

Let $\mathbb{P}(a|l) = \mathbb{P}(a \text{ out of } l \text{ neighbors infected})$. In terms of the infected link probability α, we can now specify the scaled transition probabilities[1] of the process x_n:

$$\bar{P}_{21}(l, \alpha_n) \triangleq \frac{1}{\rho(l)} \mathbb{P}\left(x_{n+1}(l) = x_n(l) + \frac{1}{M(l)}\right)$$
$$= (1 - x_n(l)) \sum_{a=0}^{l} \lambda P_{21}(l, a) \, \mathbb{P}(a|l)$$
$$= (1 - x_n(l)) \sum_{a=0}^{l} \lambda P_{21}(l, a) \binom{l}{a} \alpha_n^a (1 - \alpha_n)^{l-a},$$
$$(6)$$

$$\bar{P}_{12}(l, \alpha_n) \triangleq \frac{1}{\rho(l)} \mathbb{P}\left(x_{n+1}(l) = x_n(l) - \frac{1}{M(l)}\right)$$
$$= x_n(l) \sum_{a=0}^{l} \lambda P_{12}(l, a) \binom{l}{a} \alpha_n^a (1 - \alpha_n)^{l-a}. \quad (7)$$

where P_{12}, P_{21} are defined in (3). In (6) and (7), the notation α_n is the short form for $\alpha(x_n)$ and $\lambda > 0$ is a scaling factor referred to as the diffusion parameter.

2) Mean Field Dynamics: We are interested in modeling the evolution of infected population using mean field dynamics. An importatnt feature of the mean field dynamics model is that it has a state of dimension L compared to the intractable state dimension $\Pi_{l=1}^{L} M(l)$ of the infected degree distribution vector $\bar{x}$. Denote the unit $L-1$ dimensional simplex as $\Pi(L)$:

$$\Pi(L) \triangleq \{x \in \mathbb{R}^L : \mathbf{1}_L' x = 1, 0 \leq x(i) \leq 1 \text{ for all } i \in \{1, 2, \ldots, L\}\}$$
$$(8)$$

Note that the infected degree vector $\bar{x}_n$ is not a probability distribution as it contains the relative density of infected nodes

given the degree. However, scaling each node of degree l by $\frac{1}{\rho(l)}$, it is seen that the infected degree distribution $\bar{x}_n \in \Pi(L)$.

Theorem 1 (Mean Field Dynamics, [33])**.** *Consider the deterministic mean field dynamics process with state $\bar{x}_n \in \Pi(L)$ (the $L-1$ dimensional unit simplex):*

$$\bar{x}_{n+1}(l) = \bar{x}_n(l) + \frac{1}{M}\left[\bar{P}_{12}(l, \alpha(\bar{x}_n)) - \bar{P}_{21}(l, \alpha(\bar{x}_n))\right] \quad (9)$$

Consider the martingale representation of the Markov chain $x_n \in \Pi(L)$ (the $L-1$ dimensional unit simplex):

$$x_{n+1} = x_n + \frac{1}{M}\left[P_{12}(x) - P_{21}(x)\right] + v_n, \quad \bar{x}_0 = x_0. \quad (10)$$

where v_n is a martingale difference process with $\|v_n\|_2 \leq \frac{\Gamma}{M}$ for some positive constant Γ. Then for a time horizon of N points, the deviation between the mean field dynamics $\bar{x}_n$ in (9) and actual population distribution x_n in (10) satisfies

$$\mathbb{P}\{\max_{0 \leq n \leq N} \|\bar{x}_n - x_n\|_\infty \geq \epsilon\} \leq C_1 \exp(-C_2 \epsilon^2 M)$$

for some positive constants C_1 and C_2 providing $N = O(M)$.

The proof of Theorem 1 is given in the appendix in a simple tutorial form that uses the elementary Azuma Hoeffding inequality. Theorem 1 says that for a time horizon of N points, the deviation between the mean field dynamics $\bar{x}_n$ in (9) and actual infected distribution x_n in (10) satisfies an exponential bound.

For the purposes of this paper, the key outcome of Theorem 1 is that the mean field system $\bar{x}_n$ has polynomial dynamics. These polynomial dynamics will be exploited in Sec. III for tracking the infected degree distribution.

B. Sampling

We now consider the second component of the model, namely, observations obtained by sampling the social network. For social networks with large numbers of nodes, it is often impossible to query each node[2]. This necessitates choosing a sampling methodology in order to estimate the infected degree distribution $\bar{x}$. For the purpose of estimating the infected degree distribution $\bar{x}$, the degree distribution ρ of the entire network is assumed to be known[3]. Each sampled node is asked if it is infected or not and the reply (measurement) noted. Below, we consider two popular methods for sampling large networks:

1) Uniform Sampling: At each period n, $\gamma(l)$ individuals are sampled[4] independently and uniformly from the population $M(l)$ comprising of agents with connectivity degree l. That is, a uniform distributed i.i.d. sequence of nodes, denoted by $\{m_l, l = 1 : \gamma(l)\}$, is generated from the population $M(l)$. From these independent samples, the empirical infection

---

[1] The transition probabilities are scaled by the degree distribution ρl for notational convenience. Indeed, since $M(l) = M\rho(l)$, by using these scaled probabilities we can express the dynamics of the process x in terms of the same-step size $1/M$. We assume that the degree distribution $\rho(l)$, $l \in \{1, 2, \ldots, L\}$, is uniformly bounded away from zero.

[2] An an example, in the case of the MSM social network there is no comprehensive list of all the members of the social network and the members only respond to surveys when prompted by someone already in their network.

[3] In Sec. IV, a Bayesian filter using which the degree distribution ρ itself can be estimated is outlined.

[4] For large population sizes M, sampling with and without replacement are equivalent.

distribution $\hat{\bar{x}}_n(l)$ of degree l nodes at each time n is obtained as

$$\hat{\bar{x}}_n(l) = \frac{1}{\gamma(l)} \sum_{l=1}^{\gamma(l)} \{s_n^{(m_l)} = 1\}. \tag{11}$$

At each time n, the empirical distribution $\hat{\bar{x}}_n$ can be viewed as noisy observations of the infected distribution $\bar{x}_n$.

2) MCMC Based Respondent-Driven Sampling (RDS): Respondent-driven sampling (RDS) was introduced by Heckathorn [18], [19] as an approach for sampling from hidden populations in social networks and has gained enormous popularity in recent years. RDS is a variant of the well known method of snowball sampling where current sample members recruit future sample members. The RDS procedure is as follows: A small number of people in the target population serve as seeds. After participating in the study, the seeds recruit other people they know through the social network in the target population. The sampling continues according to this procedure with current sample members recruiting the next wave of sample members until the desired sampling size is reached. Typically, monetary compensations are provided for participating in the data collection and recruitment.

RDS can be viewed as a form of Markov Chain Monte Carlo (MCMC) sampling (see [34] for an excellent exposition). Let $\{m_l, l = 1 : \gamma(l)\}$ be the realization of an aperiodic irreducible Markov chain with state space $M(l)$ comprising of nodes of degree l. This Markov chain models the individuals of degree l that are snowball sampled, namely, the first individual m_1 is sampled and then recruits the second individual m_2 to be sampled, who then recruits m_3 and so on. Instead of the independent sample estimator (11), an asymptotically unbiased MCMC estimate is then generated as

$$\frac{\sum_{l=1}^{\gamma(l)} \frac{I(s_n^{(m_l)}=1)}{\pi(m_l)}}{\sum_{l=1}^{\gamma(l)} \frac{1}{\pi(m_l)}} \tag{12}$$

where $\pi(m)$, $m \in M(l)$, denotes the stationary distribution of the Markov chain. For example, a reversible Markov chain with prescribed stationary distribution is straightforwardly generated by the Metropolis Hastings algorithm.

In RDS, the transition matrix and, hence, the stationary distribution π in the estimator (12) is specified as follows: Assume that edges between any two nodes i and j have symmetric weights W_{ij} (i.e., $W_{ij} = W_{ji}$, equivalently, the network is undirected). In RDS, node i recruits node j with transition probability $W_{ij}/\sum_j W_{ij}$. Then, it can be easily seen that the stationary distribution is $\pi(i) = \sum_{j \in V} W_{ij}/\sum_{i \in V, j \in V} W_{ij}$. Using this stationary distribution, along with the above transition probabilities for sampling agents in (12), yields the RDS algorithm.

It is well known that a Markov chain over a non-bipartite connected undirected network is aperiodic. Then, the initial seed for the RDS algorithm can be picked arbitrarily, and the above estimator is an asymptotically unbiased estimator.

The key outcome of Sec. II-B is that by the central limit theorem (for an irreducible aperiodic finite state Markov chain), the estimate of the probability that a node is infected in a large population (given its degree) is Gaussian. Therefore, the sample observations can be expressed as

$$y_n = C\bar{x}_n + v_n \tag{13}$$

where $v_n \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ is the observation noise with the covariance matrix $\mathbf{R}$ and observation matrix C dependent on the sampling process and $\bar{x}_n \in \mathbb{R}^L$ is the polynomial function of the infected degree probabilities (9).

III. Non-linear filter for Bayesian Tracking of Infected Distribution

In Sec. II, we formulated the mean field dynamics for the infected degree distribution as a polynomial dynamical system (9) with linear Gaussian observations (13) due to sampling the network. In this section, we consider Bayesian filtering of the degree infection probabilities for large networks.

A. Optimal Filter for Polynomial Dynamics

To estimate the infected degree distribution using the sampled observations (13), we employ the optimal Bayesian filter recently developed in [21] for polynomial systems. It is shown in [21] that for Gaussian systems with polynomial dynamics, one can devise a finite dimensional filter (based on the Kalman filter) to compute the conditional mean estimate.

Rather than repeating the optimal filtering equations from [21], to save space we present the relevant terms in the model that are used in the filtering equations. Let $\hat{\bar{x}}_n^-$ and $\hat{\bar{x}}_n^+$ denote the priori and posteriori expectation of the state vector $\bar{x}_n$ and let $Y_n = \{y_{0:n}\}$ denote the observation process. Let

$$f(\bar{x}) = A_0 + A_1\bar{x} + A_2\bar{x}\bar{x}' + A_3\bar{x}\bar{x}\bar{x}' + \dots \tag{14}$$

denote the polynomial that dictates the evolution in (9). Here $A_i \in \mathbb{R}^{L \times L \times \dots \times L}$ is an $i+1$ dimensional tensor and

$$A_i\bar{x}\bar{x}\dots\bar{x}' = \sum_{j_1,j_2,j_3,\dots,j_i} A_{(:),j_1,j_2,\dots,j_i}\bar{x}_{j_1}\bar{x}_{j_2}\dots\bar{x}_{j_i}$$

The non-linear filter update equations are given as [21]:

$$\begin{aligned}
\hat{\bar{x}}_n^- &= \mathbb{E}[f(\bar{x}_{n-1})|Y_{n-1}] \\
H_n^- &= \mathbb{E}[(\bar{x}_n - \hat{\bar{x}}_n)(\bar{x}_n - \hat{\bar{x}}_n)'|Y_{n-1}] \\
K_n &= H_n^- C'(R_n + CH_n^- C')^{-1} \\
\hat{\bar{x}}_n^+ &= \hat{\bar{x}}_n^- + H_n^- C'(R_n + CH_n^- C')^{-1}(y_n - C\hat{\bar{x}}_n^-) \\
H_n^+ &= (I - K_nC)H_n^-(I - K_nC)' + K_nR_nK_n'
\end{aligned} \tag{15}$$

These estimates are computed using the higher order moments of the Gaussian conditional random variable $\bar{x}_n - \hat{\bar{x}}_n$: $\mathbb{E}[(\bar{x}_n - \hat{\bar{x}}_n)^2]) = H_n$, $\mathbb{E}[(\bar{x}_n - \hat{\bar{x}}_n)^6]) = 15H_n^3$. The filter relies upon being able to compute the expectation $\mathbb{E}[f(\bar{x}_{n-1})f'(\bar{x}_{n-1})|Y_{n-1}]$ in terms of $\hat{\bar{x}}_{n-1}$ and H_n^-, which is possible when $f(\cdot)$ is a polynomial function.

To derive the filter expressions for the infection dynamics, (9) is expanded and terms are grouped according to their order

in the state $\bar{x}_n$ to generate the tensors A_i of (14). Consider $\bar{x}_n(l)$:

$$\bar{x}_{n+1}(l) = \bar{x}_n(l) + \frac{1}{M}\left[\bar{P}_{12}(l,\alpha_n)) - \bar{P}_{21}(l,\alpha_n)\right]$$

$$= \bar{x}(l)\sum_{a=0}^{l}\lambda P_{12}(l,a)\binom{l}{a}\alpha_n^a(1-\alpha_n)^{l-a}$$

$$+ (1-\bar{x}(l))\sum_{a=0}^{l}\lambda P_{21}(l,a)\binom{l}{a}\alpha_n^a(1-\alpha_n)^{l-a} \tag{16}$$

The average degree is $\sum_l^L l\,\rho(l)$ and the link probability given in (5) can be expressed as $\alpha_n = \phi'\bar{x}_n$, where ϕ is defined as

$$\phi = \left[\frac{\rho(1)}{\sum_l^L l\,\rho(l)}, \frac{2\rho(2)}{\sum_l^L l\,\rho(l)}, \ldots, \frac{L\rho(L)}{\sum_l^L l\,\rho(l)}\right]'$$

We expand (16) into a sum of terms that are polynomial in $\bar{x}_n$. We illustrate this expansion for degree $l = 2$, noting that expansions of any other degree follow the same process. For all terms with factor P_{12} there is a corresponding term with factor P_{21}, so for convenience we will account for all of the former with Ω and the latter with Ω^*. Ω is then all the terms in an expanded (16), with $l = 2$, containing P_{12}.

$$\Omega = \lambda\Bigg[\frac{1}{M}P_{12}(2,0)(\phi'\bar{x}_n)^2 + \frac{2}{M}P_{12}(2,1)(\phi'\bar{x}_n)$$

$$- \frac{2}{M}P_{12}(2,1)(\phi'\bar{x}_n)^2 + \frac{1}{M}P_{12}(2,2)$$

$$- \frac{2}{M}P_{12}(2,2)(\phi'\bar{x}_n) + \frac{1}{M}P_{12}(2,2)(\phi'\bar{x}_n)^2$$

$$- \frac{\bar{x}_n}{M}P_{12}(2,0)(\phi'\bar{x}_n)^2 - \frac{2\bar{x}_n}{M}P_{12}(2,1)(\phi'\bar{x}_n)$$

$$+ \frac{2\bar{x}_n}{M}P_{12}(2,0)(\phi'\bar{x}_n)^2 - \frac{\bar{x}_n}{M}P_{12}(2,2)$$

$$+ \frac{2\bar{x}_n}{M}P_{12}(2,2)(\phi'\bar{x}_n) - \frac{\bar{x}_n}{M}P_{12}(2,2)(\phi'\bar{x}_n)^2\Bigg] \tag{17}$$

and so

$$\bar{x}_{n+1}(2) = \bar{x}_n(2) + \Omega + \Omega^* \tag{18}$$

By grouping all of the terms of (18) by their order in $\bar{x}_n$ we can generate the tensors of (14). The contributions to the tensors of (14) by Ω are:

$$A_0(2) = \frac{\lambda P_{12}(2,2)}{M}$$

$$A_1(2,:) = \phi\left[\frac{2\lambda(P_{12}(2,1) - P_{12}(2,0))}{M}\right]$$

$$A_2(2,:,:) = \phi\phi'\left[\frac{\lambda(P_{12}(2,0) - 2P_{12}(2,1) + P_{12}(2,2))}{M}\right]$$

$$A_2(2,2,:) = \phi\left[\frac{2\lambda(P_{12}(2,1) - P_{12}(2,0))}{M}\right]$$

$$A_3(2,2,:,:) = -\phi\phi'\left[\frac{\lambda(P_{12}(2,0) - 2P_{12}(2,1) + P_{12}(2,2))}{M}\right] \tag{19}$$

By following (19) for Ω and Ω^* for all l, we are able to generate all the coefficients in the tensors of (14) from P_{12}, P_{21}, and $\rho(l)$. We note that the polynomial that defines

the dynamics of the network is of order $L^* + 1$, where L^* is the highest degree node with complex dynamics, i.e: $P_{21}(l,a) = P_{12}(l,a) = \kappa \quad \forall \quad l > L^*$ and a, where κ is constant with respect to both l and a.

B. Posterior Cramér-Rao bounds

The lower bound for the mean square error of the optimal filter is evaluated using the well known Posterior Cramér-Rao Lower Bound (PCRLB), [22]. Below we formulate these for the polynomial dynamical system (9) and linear Gaussian observations (13) with a brief the derivation in the appendix. Since there is no state noise in (9), to compute the PCRLB, we perturb the state evolution in (9) with pairwise independent Gaussian random vectors having covariance matrix $\mathbf{Q} = \epsilon I$, replacing the singular state evolution by a perturbed system $p_\epsilon(\bar{x}_{n+1}|\bar{x}_n)$. We have,

$$-\log p_\epsilon(\bar{x}_{n+1}|\bar{x}_n) = c + \frac{1}{2}\{\bar{x}_{n+1} - f_n(\bar{x}_n)\}'\mathbf{Q}^{-1}\{x_{n+1} - f_n(\bar{x}_n)\} \tag{20}$$

Let $J(X_n)$ denote the $((n+1)L \times (n+1)L)$ Fisher information matrix[5] of X_n [22], with $X_n = [X_{n-1}, x_n]$. The following recursion is used to evaluate J_n in [22]. Let $Y_n = [Y_{n-1}, y_n]$ denote the observation sequence at time n, $\mathbb{P}(X, Y) = p_n$ denote the joint distribution and $\Delta_x^y = \nabla_x\nabla_y'$ denote the vector differential operator. Then [22],

$$J_n = \mathbb{E}\{-\Delta_{x_n}^{x_n}\log(p_n)\} -$$
$$\mathbb{E}\{-\Delta_{x_n}^{X_{n-1}}\log(p_n)\}[\mathbb{E}\{-\Delta_{X_{n-1}}^{X_{n-1}}\log(p_n)\}]^{-1}\mathbb{E}\{-\Delta_{X_{n-1}}^{x_n}\log(p_n)\}$$

and corresponds to the inverse of the $L \times L$ right lower block of $[J(X_n)]^{-1}$. The recursion for J_n is given by:

$$J_{n+1} = D_{\epsilon,n}^{22} - D_{\epsilon,n}^{21}\left(J_n + D_{\epsilon,n}^{11}\right)^{-1}D_{\epsilon,n}^{12}$$

where[6]

$$D_{\epsilon,n}^{11} = \frac{1}{\epsilon}\mathbb{E}\{(\nabla_{\bar{x}_n}f_n'(\bar{x}_n))(\nabla_{\bar{x}_n}f_n'(\bar{x}_n))'\}$$

$$D_{\epsilon,n}^{12} = \frac{1}{\epsilon}\mathbb{E}\{\nabla_{\bar{x}_n}f_n'(\bar{x}_n)\}$$

$$D_{\epsilon,n}^{21} = \{D_n^{12}\}'$$

$$D_{\epsilon,n}^{22} = \frac{1}{\epsilon}I + C\mathbf{R}^{-1}C' \tag{21}$$

where C is the linear observation matrix and $\mathbf{R}$ is the observation noise covariance matrix of (13) and $\nabla_{\bar{x}_n}f_n'(\bar{x}_n)$ is given as:

$$\nabla_{\bar{x}_n}f_n'(\bar{x}_n) = \left[\frac{\partial f}{\partial\bar{x}_n(0)}, \frac{\partial f}{\partial\bar{x}_n(1)}, \ldots, \frac{\partial f}{\partial\bar{x}_n(D)},\right]'$$

[5]The error covariance $\mathscr{P}$ is,

$$\mathscr{P} = \mathbb{E}\{(\hat{X} - X)(\hat{X} - X)'\} \geq J^{-1}$$

where J is the Fisher Information matrix, X is the state, $\hat{X}$ is the state estimate, and Y measured data. The elements of the Fisher information matrix J are given by:

$$J_{ij} = \mathbb{E}\left(\frac{\partial^2\log p_{x,y}(X,Y)}{\partial X_i\partial X_j}\right)$$

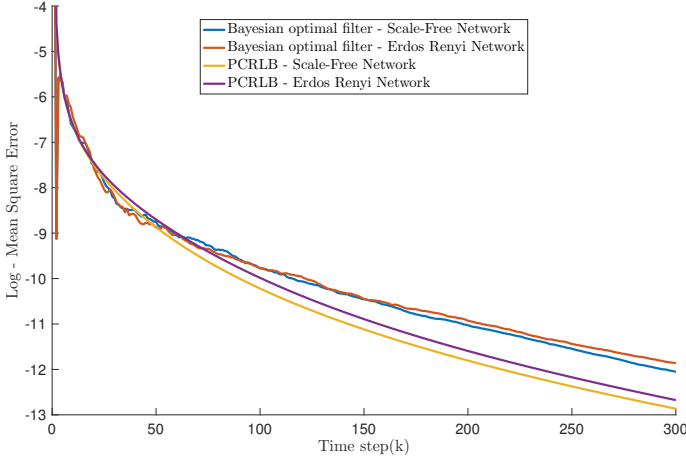[6]see Appendix for a brief derivation

Fig. 1: Mean square error and PCRLB for two different networks - Scale-free and Erdős-Rényi. It can be seen that the filter performs near optimally and has no discernible dependence on the degree distribution. Both PCRLB and its slope are insensitive to the underlying distribution.

$$\frac{\partial f}{\partial \bar{x}_n(0)} = 0 + \frac{\partial}{\partial \bar{x}_n(0)} \left[A_1 \bar{x}_n \right] + \frac{\partial}{\partial \bar{x}_n(0)} \left[A_2 \bar{x}_n \bar{x}'_n \right] + \frac{\partial}{\partial \bar{x}_n(0)} \left[A_3 \bar{x}_n \bar{x}_n \bar{x}'_n \right] \tag{22}$$

Thus

$$\nabla_{\bar{x}_n} f'_n(\bar{x}_n) = A_1 + (A_2 + A'_2)\bar{x}_n + (A_{3_{ijk}} + A_{3_{jki}} + A_{3_{kji}})\bar{x}_n \bar{x}'_n \tag{23}$$

where $A_{3_{ijk}}$ indicates the ordering of indices of the tensor A_3 and is analogous to a higher dimensional transpose.

PCRLB: Erdős-Rényi vs Scale Free: We used the above method to compute the PCRLB for the mean field dynamics model (9), with linear Gaussian observations (13). We first consider a scale-free network with degree distribution $\rho(l) \propto l^{-\gamma}$. Such networks arise in online social networks such as Twitter [31] and in the link network of the World Wide Web [30]. The second network we consider is an Erdős-Rényi network with degree distribution $\rho(l) \propto \frac{e^{-\lambda l}}{l!}$. Fig.5 shows the PCRLB for a scale-free network with $\gamma = 2.7$ and an Erdős Rényi network with $\lambda = 2.7$. The value $\lambda = 2.7$ was chosen since it is similar to the the out-degree of the World Wide Web, see [35]. For both networks, the infection transition probabilities were a random stochastic matrix. The displayed mean square errors are the average of 100 independent simulations.

Interestingly, it can be seen from Fig. 1 that both PCRLB and its slope are insensitive to the underlying network structure when observation noise variance, $\mathbf{R}$ in (13), is not network dependent. In Sec. V, we will show that the performance of the optimal filter is also insensitive to the underlying network structure. These suggest that for tracking the infected degree distribution, precise knowledge of the underlying network distribution is not required.

IV. ANALYSIS OF INFECTION DIFFUSION IN EVOLVING SOCIAL NETWORKS

So far in this paper, we have discussed estimating infection diffusion in a fixed network. In this section, we consider social networks that evolve with time, represented by changing degree distributions, and analyse their effect on the diffusion of infection over time. Since information diffusion occurs at a faster time scale compared to forming connections in social networks, we consider a two time scale formulation: where the degree distribution of the underlying network changes on a *slow time* scale and the infection diffuses over a *faster time* scale. There are various generative models for large real world networks in the literature, see [26], [27], and the references therein. In this paper, we consider the preferential attachment model discussed extensively in [26], to model the time evolution of the underlying degree distribution. The primary motivation for choosing a preferential attachment graph is that it is the simplest graph whose steady state distribution obeys a power law [26], which commonly arises in most real world networks, see [28], [29], [30].

A. Preferential Attachment Model for Network Evolution

A network evolving according to the preferential attachment model[7] is characterized by two parameters - a probability p and an initial graph G_0. The graph evolves as follows:

1) *Vertex-Step*: A new vertex is added and is connected to a vertex of the existing graph chosen independently with probability proportional to its degree.
2) *Edge-Step*: A new edge is added between two vertices of the graph chosen independently with probability proportional to their degrees.

At each time step, with probability p, Vertex-step is realized and with probability $1 - p$, Edge-step is realized.

We are interested in how the diffusion thresholds (see Def. 1 below) change when the graph is evolving with different parameters. In order to determine how the diffusion threshold depends on the graph evolution, we first need to specify a generative model for the evolution.

Let $m_{d,k}$ denote the number of vertices of degree d at time k and

$$\rho_{d,k} = \frac{\mathbb{E}(m_{d,k})}{M_k}$$

denote the expected fraction (degree distribution) of vertices of degree d at time k, where M_k denotes the total number of nodes or agents at time k in the network. A vertex of degree d at time k could have come from two cases, either it was a vertex of degree d at time $k - 1$ and had no edge added to it, or it was a vertex of degree $d - 1$ at time $k - 1$ and a new edge was put in adjacent to it. The recursion for the degree distribution $\rho_{d,k}$ can be expressed as [26]:

$$\rho_{d,k} = \left(1 - \frac{(2-p)d}{2k} \right) \rho_{d,k-1} + \left(\frac{(d-1)(2-p)}{2k} \right) \rho_{d-1,k-1} \tag{24}$$

[7] The structural results we present here apply to any dynamical graph model that satisfies the stochastic dominance conditions given in Theorem 2.

Let $\rho_k = [\rho_{1,k}, \rho_{2,k}, \ldots, \rho_{N,k}, \rho_{N(+),k}]$ denote the degree distribution at time k, with $\rho_{N(+),k}$ representing all degrees greater than N. The grouping of all states into one compound state is for notational convenience and as will be shown below, it is amenable to analysis. The compound state is modeled as an absorbing state as either edges or vertices are added during network evolution and no deletion takes place - once a node is of degree greater than d, it will continue to have degree atleast d. In matrix form, the recursion in equation (25) can be written as

$$\rho_k = H'_k \rho_{k-1} \tag{25}$$

where the matrix H_k can be expressed as $H_k = I + \epsilon_k F$ and

$$F = \begin{bmatrix} -(2-p) & (2-p) & 0 & \ldots 0 & 0 \\ 0 & -(2(2-p)) & (2(2-p)) & \ldots 0 & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & & \ldots & 0 & 0 \end{bmatrix}$$

where F is a generator matrix for the graph evolution having row sum equal to 0 and $\epsilon_k = \frac{1}{k}$. Clearly, the matrix H is a stochastic matrix with row sum equal to 1. In what follows, we will give sufficient conditions to compare the evolution dynamics of two social networks when the underlying graphs are changing according to a preferential attachment scheme.

On networks having fixed degree distribution (fixed networks), [1] identified conditions under which a network is susceptible to an epidemic using a mean-field approach and provided a closed form solution for the diffusion threshold for infection diffusion[8]. It was shown that under reasonable conditions on the infection probabilities, the diffusion threshold decreases with the mean preserving spread. In this paper, we extend the analysis of diffusion thresholds to evolving networks by providing sufficient conditions on the parameters that dictate the evolution.

B. Effect on Diffusion Threshold in SIS model

In this section, we establish a relation between the addition probability p in the preferential attachment model and the diffusion threshold λ_* in the SIS model.

Definition 1 ([1]). *The Diffusion Threshold (λ_*) is*

$$\lambda_* = \inf\{\lambda > 0 : \bar{x}_\infty \in \mathbb{R}_+^L\}$$

where $\bar{x}_\infty$ denotes the asymptotic infected degree distribution in (9).

In words, diffusion threshold $\lambda_* \in \mathbb{R}^+$ is a value of λ in (6) and (7), such that starting from a small fraction of infected agents in the network, the dynamics converges to a *positive* fraction of infected agents for all $\lambda > \lambda_*$. The existence of asymptotic infected degree distribution $\bar{x}_\infty$ is given by Lemma 1. For a proof, the reader is referred to [1].

[8]

$$\lambda_* = \frac{\sum_l^L l\,\rho(l)}{\sum_{l \geq 1} l^2 \rho(l) P_{ij}(l,1)}, \quad \text{where } P_{ij}(l,a) \text{ was defined in (3).}$$

Let $Q(\alpha) = \frac{1}{l}\sum_{l \geq 1} l\rho(l)\frac{\bar{P}_{21}(l,\alpha)}{(1-x(l))+\bar{P}_{21}(l,\alpha)}$, where $\bar{P}_{21}(l,\alpha)$ and $(1-x(l))$ are defined in (6).

Lemma 1 ([1]). $\bar{x}_\infty$ exists iff $\frac{dQ(0)}{d\alpha} > 1$.

In words, there exists an asymptotic infected degree distribution if $Q(\alpha)$ has a slope greater than 45^0 at the origin. The asymptotic infected link probability α_∞ can be calculated from $\bar{x}_\infty$ using (5).

The following theorem characterizes of the diffusion threshold of the SIS model as a function of the addition probability p[9], for a preferential attachment graph.

Theorem 2. *Consider a time evolving preferential attachment network with addition probability $p > 0$. For any initial degree distribution ρ_0, let $\rho_k(p)$ denote the degree distribution at time k and $\lambda_*^p(k)$ denote the diffusion threshold for the network with addition probability p. Then,*
  1) *$\rho_k(p)$ is first-order stochastically decreasing in p for every $k > 1$, where k denotes the slow-time index.*
  2) *$\lambda_*^p(k)$ is increasing in p.*

The proof of Theorem 2 is given in the appendix. The first part of Theorem 2 asserts that $\rho_k(p_2) >_{sd} \rho_k(p_1)$, [10] for $p_1 > p_2$, i.e, networks that have higher probability of edge addition always have higher degree distributions as the network evolves. The second part of Theorem 2 asserts that the diffusion threshold increases with the probability p of adding new vertices. This can be interpreted as networks with smaller number of nodes of higher degree requiring a larger fraction of infected individuals to have a positive fraction of infected individuals when the dynamics converges.

Remark: The key result above is the stochastic dominance structure for the preferential attachment graph - it is of interest in future work to give sufficient conditions on more general types of dynamic graph models.

C. Filtering for Estimating the degree distribution

So far in this paper, it was assumed that the degree distribution of the social network is known. In this section, we outline a Bayesian filter to estimate the degree distribution, when it evolves according to the preferential attachment model[11].

In Sec. IV-B, it was shown that the social network structure (degree distribution ρ) plays a significant role in the asymptotic infected degree distribution (Lemma 1). However, since the infection diffusion occurs at a faster time scale (n) compared to forming connections in social networks (time scale k), the asymptotic distribution $\bar{x}_\infty$ can in turn influence the network rearrangement at a future time $k + 1$. For the preferential attachment model of Sec. IV-A, this influence can be modeled as the probability p being dependent on α_∞ (which depends on $\bar{x}_\infty$).

---

[9]It should be noted that the probability p itself can be a function of α_∞ as the degree distribution and infected degree distribution are evolving on different time scales.

[10]$>_{sd}$ denotes first-order stochastic dominance (see Appendix for definition)

[11]Recall that Sec III deals with filtering to track the infected degree distribution in a fixed network.

It is interesting to note that the evolution of degree distribution in (25) has the form of a Chapman-Kolmogorov equation for a Markov chain η having the state space $\{1, 2, \ldots, N^+\}$. In other words, Chapman-Kolmogorov equation is a generative model for the evolution of the network. We exploit this property to estimate the degree distribution by deriving a representative sample that captures the link between the degree distribution ρ and the asymptotic degree distribution $\bar{x}_\infty$. This link could be an important factor in determining the way connections are formed in social networks; see for example, [36]; where, similarity between individuals (Homophily) breeds connection. Nodes which are not previously connected, but being infected currently increases the probability of forming a link at a future time instant.

Let $k = 0, 1, \ldots$ denote the slow time index. Let z_k[12] denote the observation at time k. Below, we consider the mode of the asymptotic infected degree distribution $\bar{x}_\infty \in \mathbb{R}^L$ as the representative sample z_k at time k. The mode of the infected degree distribution gives the degree with the largest fraction of infected individuals and tracking the mode can provide useful information on the nature of infection diffusion over time k. Let the initial estimate be $\hat{\rho}_0$, which denotes the probability distribution[13] of the mode (of asymptotic infected degree distribution) over the set $\{1, 2, \ldots, N^+\}$.

Given this observation z_k at time k, and the dynamics of the degree distribution (IV-A), define the posterior distribution of the degree distribution as

$$\hat{\rho}_{k+1} = \mathbb{P}(\rho_{k+1} | z_1, \ldots, z_{k+1})$$

Then it is easily seen that the evolution of the posterior distribution is given as the Hidden Markov Model (HMM) filter [33]:

$$\hat{\rho}_{k+1} = \frac{B_{z_k} H_k' \hat{\rho}_k}{\mathbf{1}' B_{z_k} H_k' \hat{\rho}_k} \tag{26}$$

where $B_{z_k} = \mathrm{diag}(\mathbb{P}(z_k | \eta_k = i))$ for $i \in \{1, 2, \ldots, N^+\}$ is a diagonal matrix.

It should be noted that the at time $k + 1$, both the estimate of the mode and the degree distribution $\hat{\rho}_{k+1}$ are obtained.

To summarize, (26) together with the filter (Sec. III-A) constitutes a two time scale tracking algorithm: on the slow time scale, the degree distribution of the social network is updated based on sampling according to (26). On the fast time scale, these estimates are used in the filter (Sec. III-A) to track the infected degree distribution. We refer to [37] for a formal proof of the optimality of this time two-time scale filtering algorithm.

V. NUMERICAL EXAMPLES AND TWITTER DATA

This section presents two main results. First, computer simulations are used to illustrate the performance of the filtering algorithm for tracking the evolving infected degree distribution. The sensitivity of the filter to the underlying network (Erdős-Rényi vs Scale Free) is examined.[14] Second,

the SIS model is illustrated using a real Twitter dataset and the infection diffusion is tracked using the non-linear filter in Sec. III. It is shown using a goodness of fit test that the SIS model yields a satisfactory fit to the Twitter data and also that the Bayesian filter performs satisfactorily.

A. Filtering on Erdos-Renyi vs Scale-Free Networks

Recall a scale-free (SF) network has a degree distribution $\rho(l) \propto l^{-\gamma}$ while an Erdős-Rényi (ER) network has a degree distribution $\rho(l) \propto \frac{e^{-\lambda l}}{l!}$. Given the transition probabilities, degree distribution, and initial conditions, we can simulate trajectories of the mean-field dynamics (9) and generate observations according to (11). Below, we illustrate the effect of sampling on filtering.

1) Effect of Sampling on Filtering: Observation noise variance $\mathbf{R}$ in (13) depends on the sampling method employed. Through this dependence, if sampling is network dependent, so is the observation noise variance, and transitively the estimate is network dependent. The effect of sampling on filtering is illustrated using Uniform sampling of Sec. II-B.

Uniform Sampling: A simulated diffusion state and observation trajectory are shown in Fig. 2, and the respective mean square filter error is shown in Fig. 3.[15] The transition probabilities are a random stochastic matrix (same for both ER and SF networks); and observations are simulated according to (13) with $C = I$, $x_0 = \frac{1}{2} \forall l$, and constant observation noise variance $\mathbf{R} = (5 \times 10^{-3}) I$, where I is the identity matrix. These parameters were chosen arbitrarily. It is observed that for constant observation noise variance, there is no discernible difference between Scale-Free and Erdős-Rényi networks in the accuracy of the filtered state estimates.
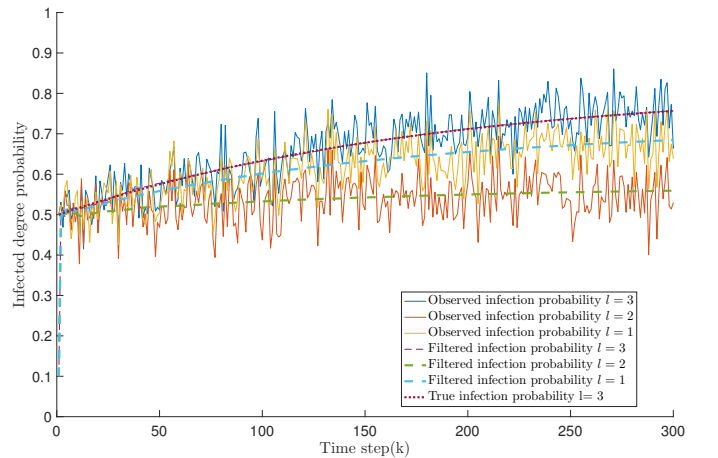

Fig. 2: Diffusion of infection probability trajectories and their corresponding filtered states in a scale-free network. It can be seen that the estimates converge to the true state for all degrees.

In Fig. 4, the transition probabilities are a random stochastic matrix; and observations are simulated according to (13) with $C = I$ $x_0 = \frac{1}{2} \forall l$ and a non-diagonal *random* observation noise variance matrix $\mathbf{R}$. We observe that distributions that are more uniformly spread amongst all degrees result in fewer

---

[12]Observation z_k can be a scalar or a vector depending on the application.

[13]Note that here the degree distribution $\hat{\rho}$ is interpreted as the probability distribution of the mode of asymptotic infected degree distribution $\bar{x}_\infty$.

[14]Recall Sec. III-B examined the sensitivity of the posterior Cramér-Rao bounds to the underlying network.

[15]The performance of the moving average filter is provided for comparison.
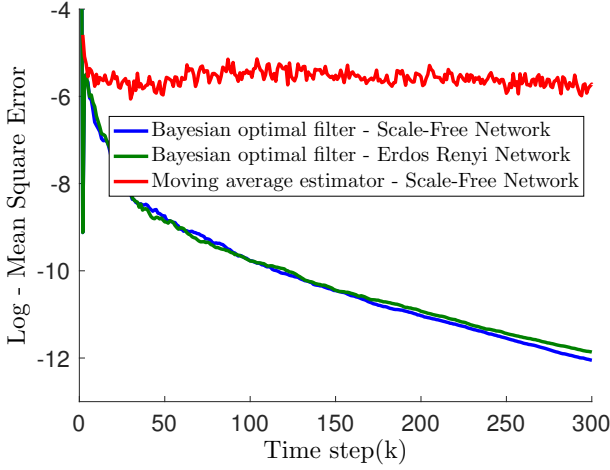
Fig. 3: Mean Squared Error for both Erdős-Rényi and scale-free networks are compared. For both networks, the same transition probabilities and initial conditions are used. The averages shown have been averaged over 50 simulations. It is observed that for constant observation noise variance, the MSE for both Scale-Free and Erdős-Rényi networks are indistinguishable.
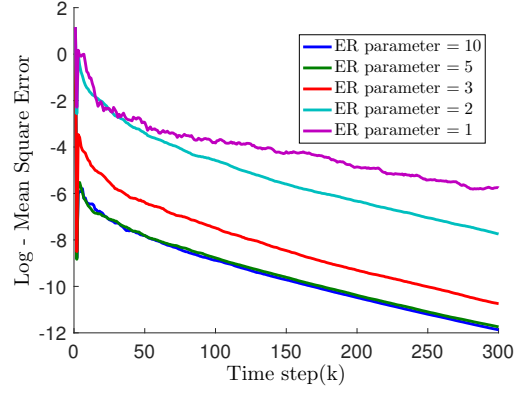
degrees with high observation noise variance, which in turn reduces the total mean square error. In scale-free networks, as shown in Fig. 4b, for larger γ, the degree distribution decays more quickly, thus there is less probability mass on nodes of higher degree and they are less frequently sampled. According to (11), fewer samples corresponds to a higher variance and therefore a worse estimate. In Erdős Rényi networks Fig. 4a, most of the probability mass is centered around the mean of the degree distribution and so there are fewer degrees with large observation noise variance. It can be seen that ER networks having a higher average degree λ have smaller MSE.

2) Sensitivity of Filter Performance to Mis-specified Model: We call a degree distribution mis-specified if it does not match the degree distribution of the true network. A Bayesian filter, derived with a scale-free degree distribution; a mis-specified filter, same Bayesian filter derived with the degree distribution $\rho(l) = \frac{1}{L} \quad \forall l$; and an autoregressive moving average filter with parameters computed by multivariate least square estimation (for a comprehensive review of vector auto regression and LS parameter estimation see [38]) are used to analyze the sensitivity of filter performance to mis-specified models. It is observed in Fig. 5 that, even when the degree distribution is mis-specified, the Bayesian filters outperform the moving average filter with an MSE of the order of 10^{-8}, compared to 10^{-6} of the moving average filter.
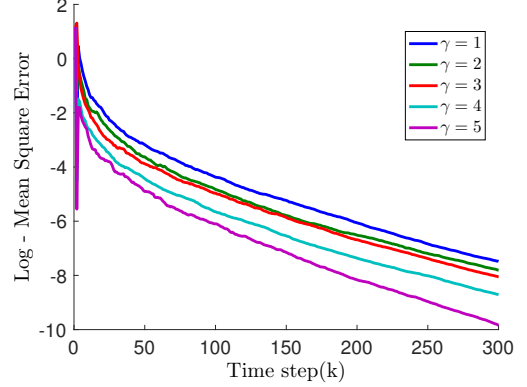
B. Analysis and Validation on Twitter Dataset

This section illustrates the tracking of infection diffusion on social networks using real diffusion data from the microblog platform Twitter. Twitter is a social media platform over which users can communicate in short < 140 character messages, images and video files. We analyze the diffusion of information through the Twitter Social network to demonstrate the effectiveness of the SIS model of Sec. II.
Twitter played a critical role in the Egyptian revolution of 2011

(a) Log mean square error in Erdős Rényi networks for varying ER parameter, λ. It can be seen the MSE decreases as λ increases.

(b) Log mean square error in scale-free networks for varying γ. For larger scale-free parameter γ it is observed that the MSE increases.

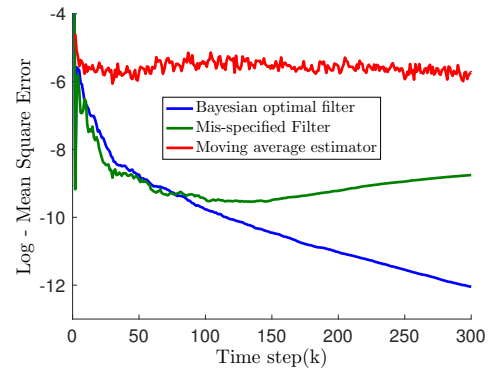**Fig. 4:** Mean square error for varying network parameters and network types.

Fig. 5: Comparison of the mean square error between a Bayesian filter, a mis-specified Bayesian filter and a vector autoregressive estimator.

or January 25th (#Jan25) uprising [39]. Twitter was used by protesters to organize the protest and recruit members and as a medium to discuss and share information about the protest. This uprising precipitated quickly and was violent, both of which acted as substantial barriers towards traditional media coverage. Below, we refer to the interest and engagement with the news of the uprising as *infection* and track the distribution of infection over time.

1) Dataset: The dataset consists of tweets sampled between January 23rd and February 8th, 2011 and are available from Twitter (http://trec.nist.gov/data/tweets/). The tweet collection period encapsulates the time-frame of the first major developments relating to the January 25th uprising event. In Twitter, a "hashtag" follows the discussion topics, i.e., a word or a phrase prefixed with the number sign #. We make use of the hashtags to track the spreading of a specific topic on Twitter. The most used hashtag related to this protest is "#Jan25". To obtain the information spreading among users participating in this protest, we filtered 26,313 tweets containing "#Jan25" published by 13,341 different users, from around 10 million tweets. These tweets contain the event of interest and the social network is (re-)constructed from them as follows: two users are connected if one user has mentioned another user ("@username") in the tweet containing "#Jan25" at least once over the duration of interest. On this constructed social network, information diffusion is analysed.

All users in the constructed social network are assumed to be susceptible initially. Users who initiate tweets on the event of interest are assumed to be infected and act as seeds for the spread of information. Once a user, say User#A becomes infected, it has some constant probability, say δ, of becoming susceptible in each time period. This modeling assumption is motivated by the frequently observed poisson-like decay of an individual's interest in social media topics [40].

2) SIS model for Twitter data: In case of the spread of engagement in Twitter, individuals can either be inactive or active depending on if they are willing and able to spread information and interest on a topic. Active users can be considered 'infected', and inactive users can become 'infected' by interacting with other 'infected' individuals, in particular, any of its active neighbours in a social network. In this way, engagement and knowledge of a topic spreads throughout the network. People can also become disinterested in a subject they have already been exposed to, in this way they are not currently engaged, but may become engaged if contacted by an infected neighbour; thus inactive individuals are assumed to be susceptible. This is the basis of the susceptible-infected-susceptible (SIS) model and its application to the diffusion of information on Twitter has been noted and studied; see [31], [32].

We adopt the SIS model for the Twitter data, mapping engagement in the January 25th uprising as an infection spreading over the Twitter network created from Twitter users. Below we analyze the goodness-of-fit of the SIS model for the Twitter data using a standard statistical test, the Kolmogorov-Smirnov test [41], and measure the square and absolute difference between the data and model predictions. The SIS model used here for evaluation utilizes empirically determined parameters as outlined in Sec. V-B3.

Model Evaluation (Goodness of Fit for SIS model): We used the Kolmogorov-Smirnov (KS) test on the empirical infected degree distribution at the final timepoint to evaluate the SIS model. The KS test statistic was 0.2286 with p-value 0.2813. The null hypothesis for this statistical test is that both the observed Twitter and SIS model infected degree distributions are samples of the same infected degree distribution. At a

confidence level of 0.01, we do not reject the null hypothesis. We also calculate the average and maximum square difference between the Twitter data and predicted SIS degree infection probabilities. The differences are calculated at each timepoint; both the averaged-over-time and maximum differences are shown. These values can be seen in Table I and the trajectories are shown in Fig. 7.

TABLE I: Goodness of fit for the SIS model: The average and maximum deviations between the Twitter data and SIS model predictions are presented. The large absolute difference in high degree nodes arises when there are few nodes of this large degree and thus the empirical infection probability deviates from the asymptotic case.

Degree	1	2	3+
Average Square Difference	0.0011	0.0014	0.0235
Average Absolute Difference	0.0273	0.0294	0.1000
Maximum Absolute Difference	0.0644	0.0719	0.8403

The low magnitude of the model deviations in Table I for the Twitter dataset and the failure to reject the hypothesis that the Twitter data and model infected degree distributions come from the same distribution, suggest that the SIS model is a satisfactory model with respect to the infection dynamics of interest in the January 25th uprising.

3) Sampling for tracking the infected degree distribution: The mean field dynamics for the SIS model can be used to track and predict the evolution of the infection on Twitter. We must generate estimates of P_{12}, P_{21}, and determine the degree distribution from samples obtained from (13). P_{12} is given by δ, since all infected nodes become susceptible with probability δ at each time point. We compute the empirical transmission rates $\hat{P}_{21}$ directly by observing the frequency with which an infected individual with l neighbors, a of which are infected, becomes infected.

$$\hat{P}_{12}(l,a) = \frac{\sum_{n=0}^{T}\sum_{m=1}^{M}\left(s_{n+1}^{(m)}=1 | s_n^{(m)}=2, D^{(m)}=l, F_n^{(m)}=a\right)}{\sum_{n=0}^{T}\sum_{m=1}^{M}\left(s_n^{(m)}=2, D^{(m)}=l, F_n^{(m)}=a\right)}$$
(27)

The degree distribution used is the empirical degree distribution, shown in Fig. 6. We analyze the degree distribution and find that it fits a power law distribution with power law exponent -2.425 which matches very closely with the power law distribution with exponent -2.412 found for the Twitter network in [42]. The true degree infection probabilities are computed directly from the entire network for each 1 minute time interval.

Next, we sample only a subset of the data using the RDS sampling scheme described in Sec. II-B, every 1 minute and track the infected degree distribution over time using the non-linear Bayesian Filtering technique described in Sec. III-A. The parameters used in the Bayesian filter are: empirical $\hat{P}_{12}$ and empirical ρ and the filter estimates are shown in Fig. 8. It is seen that the filtered estimates track the true infected degree distribution satisfactorily over time.
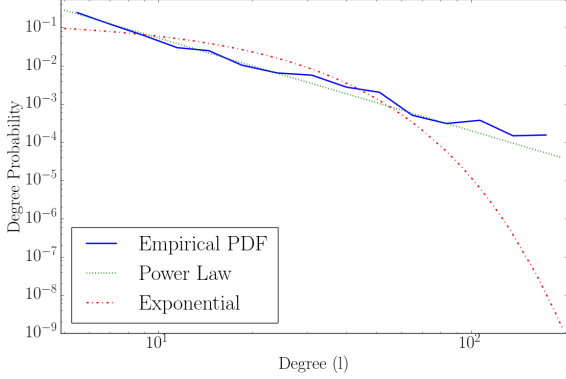
Fig. 6: The degree distribution of the constructed Twitter #Jan25 social network is analyzed and found to follow a power law distribution. The exponent parameter -2.425 of the power law distribution was chosen as the maximum likelihood estimate. The MLE and likelihood ratio were computed by the computational package in [43] according to the methods in [44]. The loglikelihood ratio between a power law and exponential distributions is 8.631, which is significant evidence that the data follows a power law distribution rather than an exponential distribution.
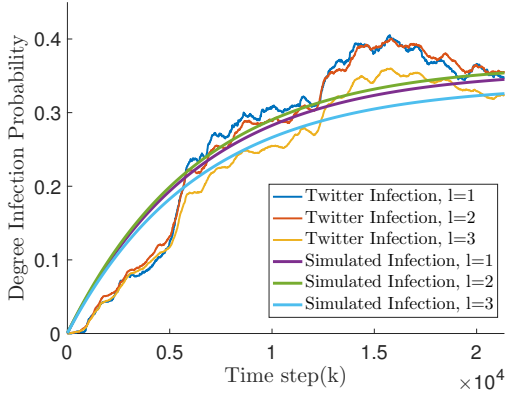

Fig. 7: The true Twitter infection and simulated Twitter infections are compared for nodes of degree $l = 0, 1, 2$. The simulated trajectories use the empirically generated ρ, δ and $\hat{P}_{12}$ (27).
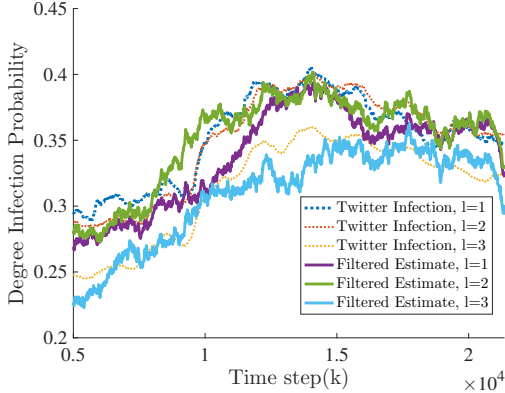

Fig. 8: The true Twitter infection and the filter estimates of the Twitter degree infection probabilities are compared. Samples are generated by RDS sampling on the #Jan25 social network. At each timestep a 10,000 node walk is performed and from this walk an observation of the infected degree distribution is generated.

VI. CONCLUSION

We considered the problem of tracking infection diffusion over large social networks by modeling the diffusion process using SIS model. The infection distribution was approximated using mean field dynamics which resulted in a state space model with polynomial dynamics. Posterior Cramér-Rao bounds were computed for the mean field dynamics and it was shown that these bounds are relatively insensitive to the type of underlying social network (Erdős-Rényi vs Scale Free network). Next, to account for the time-varying nature of the degree distribution of large real-world networks, the relationship between diffusion thresholds and the changing degree distribution was analyzed using a generative model for a network generated using the preferential attachment model. It was found that networks on which more edges are added relative to nodes, have lower diffusion thresholds. Finally, a Twitter dataset was used to illustrate how information diffusion on the Twitter platform can be modeled by a mean field dynamical SIS model, and that, under this model, we can filter and track the evolution of the degree infection probabilities over time.

APPENDIX A
PROOF OF MEAN FIELD THEOREM

The proof of the mean field dynamics approximation is given in [45], but the presentation is not readily accessible. We show below that the proof is a simple consequence of Azuma-Hoeffding inequality. A bound on the deviation between the mean field dynamics $\bar{x}_n$ in (9) and actual population distribution x_n in (10) is calculated in the form of two lemmas, Lemma 2 and Lemma 3. We first state the Azuma-Hoeffding Inequality (Theorem 3 below) which gives a bound on the deviation of a random variable from some value for the values of martingales that have bounded differences.

Theorem 3 (Azuma-Hoeffding Inequality). *Suppose* $S_N = \sum_{k=1}^{N} v_k + S_0$ *where* $\{v_k\}$ *is a martingale difference process with bounded differences satisfying* $|v_k| \leq \Delta_k$ *almost surely where* Δ_k *are finite constants. Then for any* $\epsilon > 0$,

$$\mathbb{P}(|S_N - S_0| \geq \epsilon) \leq 2 \exp\left(-\frac{\epsilon^2}{\sum_{k=1}^{N} \Delta_k^2}\right)$$

□

Define

$$\tilde{x}_n = x_k - \bar{x}_k, \quad S_N = \max_{1 \leq n \leq N} \|\sum_{k=1}^{n} v_k\|_\infty$$

Here, v_k is an L-dimensional finite-state martingale increment process with $\|v_k\|_2 \leq \frac{\Gamma}{M}$ for some positive constant Γ.

Lemma 2.

$$\|\tilde{x}_{n+1}\|_\infty \leq \|\tilde{x}_0\|_\infty + \frac{\beta}{M} \sum_{k=1}^{n} \|\tilde{x}_k\|_\infty + S_N.$$

Proof of Lemma 2: Recall $\tilde{x}_n = x_n - \bar{x}_n$. Let $H(x)$ be a Lipschitz function. Clearly,

$$\tilde{x}_{n+1} = \tilde{x}_n + \frac{1}{M}[H(x_n) - H(\bar{x}_n)] + v_n$$

$$= \tilde{x}_0 + \frac{1}{M}\sum_{k=1}^{n}[H(x_k) - H(\bar{x}_k)] + \sum_{k=1}^{n} v_k$$

$$\|\tilde{x}_{n+1}\|_\infty \leq \|\tilde{x}_0\|_\infty + \frac{1}{M}\sum_{k=1}^{n}\|[H(x_k) - H(\bar{x}_k)]\|_\infty + \|\sum_{k=1}^{n} v_k\|_\infty$$

$$\leq \|\tilde{x}_0\|_\infty + \frac{\beta}{M}\sum_{k=1}^{n}\|\tilde{x}_k\|_\infty + S_N$$

since $\|[H(x_k) - H(\bar{x}_k)]\|_\infty \leq \beta\|x_k - \bar{x}_k\|_\infty$ where β is the Lipschitz constant. $\square$

Lemma 3.

$$\mathbb{P}(S_N \geq \epsilon) \leq 2\exp\left(-\frac{\epsilon^2 M^2}{2\Gamma N}\right)$$

Proof of Lemma 3: $\|\sum_{k=1}^{n} v_k\|_\infty = \max_i |\sum_{k=1}^{n} e_i' v_k| = |\sum_{k=1}^{n} e_{i^*}' v_k|$ for some i^*. Since $e_{i^*}' v_k$ is a martingale difference process with $|e_{i^*}' v_k| \leq \sqrt{\Gamma}/M$ applying the Azuma-Hoeffding inequality (Theorem 3) yields

$$\mathbb{P}(\|\sum_{k=1}^{n} v_k\|_\infty \geq \epsilon) = \mathbb{P}(|\sum_{k=1}^{n} e_{i^*}' v_k| \geq \epsilon) \leq 2\exp\left[-\frac{\epsilon^2 M^2}{2\Gamma n}\right]$$

The right hand side is increasing with n. So clearly

$$\mathbb{P}(\max_{1\leq n\leq N}\|\sum_{k=1}^{n} v_k\|_\infty \geq \epsilon) \leq 2\exp\left[-\frac{\epsilon^2 M^2}{2\Gamma N}\right]$$

$\square$

Using Lemmas 2 and 3 the proof of Theorem 1 is as follows. Applying Gronwall's inequality[16] to Lemma 2 yields $\|\tilde{x}_n\|_\infty \leq S_N \exp\left[\frac{\beta n}{M}\right]$, which in turn implies that

$$\max_{1\leq n\leq N}\|\tilde{x}_n\|_\infty \leq S_N \exp\left[\frac{\beta N}{M}\right].$$

As a result

$$\mathbb{P}(\max_{1\leq n\leq N}\|\tilde{x}_n\|_\infty > \epsilon) \leq \mathbb{P}(S_N \exp\left[\frac{\beta N}{M}\right] > \epsilon)$$

$$= \mathbb{P}(S_N > \exp\left[-\frac{\beta N}{M}\right]\epsilon)$$

Next applying Lemma 3 to the right hand side yields

$$\mathbb{P}(\max_{1\leq n\leq N}\|\tilde{x}_n\|_\infty > \epsilon) \leq 2\exp\left(-\exp(\frac{-2\beta N}{M})\epsilon^2\frac{M^2}{2\Gamma N}\right).$$

Finally choosing $N = c_1 M$, for some positive constant c_1 yields

$$\mathbb{P}(\max_{1\leq n\leq N}\|\tilde{x}_n\|_\infty > \epsilon) \leq 2\exp(-C_2\epsilon^2 M)$$

where $C_2 = \exp(-2\beta c_1)\frac{1}{2\Gamma c_1}$. This completes the proof of Theorem 1. $\square$

[16]Gronwall's inequality: if $\{x_k\}$ and $\{b_k\}$ are non-negative sequences and $a \geq 0$, then

$$x_n \leq a + \sum_{k=1}^{n-1} x_k b_k \implies x_n \leq a\exp(\sum_{k=1}^{n-1} b_k)$$

APPENDIX B
RECURSION FOR POSTERIOR CRAMER RAO LOWER BOUND [22]

Consider a non-linear state space model given by:

$$x_{n+1} = f(x_n) + w_n$$
$$y_n = h(x_n) + v_n$$

with $w_n \sim \mathcal{N}(0, \mathbf{Q}_n)$ and $v_n \sim \mathcal{N}(0, \mathbf{R}_n)$. Then, the recursive equations for parameters estimation are given by [22]:

$$J_{n+1} = D_n^{22} - D_n^{21}\left(J_n + D_n^{11}\right)^{-1}D_n^{12}$$

where

$$D_n^{11} = \mathbb{E}\{-\Delta_{x_n}^{x_n}\log p(x_{n+1}|x_n)\}$$
$$D_n^{12} = \mathbb{E}\{-\Delta_{x_n}^{x_{n+1}}\log p(x_{n+1}|x_n)\}$$
$$D_n^{21} = \mathbb{E}\{-\Delta_{x_{n+1}}^{x_n}\log p(x_{n+1}|x_n)\} \quad (28)$$
$$D_n^{22} = \mathbb{E}\{-\Delta_{x_{n+1}}^{x_{n+1}}\log p(x_{n+1}|x_n)\}$$
$$+ \mathbb{E}\{-\Delta_{x_{n+1}}^{x_{n+1}}\log p(y_{n+1}|x_{n+1})\}$$

As was stated in Sec. II, the state evolution has polynomial dynamics $f(\cdot)$ and sampling results in linear observations C. Since our model has Gaussian state (20) and observation noise (13), we can compute the components of (28) in terms of the state and observation functions $f(\cdot)$ and C.

$$-\log p(x_{n+1}|x_n) =$$
$$c + \frac{1}{2}\{x_{n+1} - f_n(x_n)\}'\mathbf{Q}_n^{-1}\{x_{n+1} - f_n(x_n)\}$$
$$-\log p(y_{n+1}|x_{n+1}) = \quad (29)$$
$$c + \frac{1}{2}\{y_{n+1} - Cx_{n+1}\}'\mathbf{R}_n^{-1}\{y_{n+1} - Cx_{n+1}\}$$

Thus $D_n^{11}, D_n^{21}, D_n^{12}, D_n^{22}$ are given as:

$$D_n^{11} = \mathbb{E}\{(\Delta_{x_n}f_n'(x_n))\mathbf{Q}_n^{-1}(\Delta_{x_n}f_n'(x_n))'\}$$
$$D_n^{12} = \mathbb{E}\{\Delta_{x_n}f_n'(x_n)\}\mathbf{Q}_n^{-1}$$
$$D_n^{21} = \{D_n^{12}\}' \quad (30)$$
$$D_n^{22} = \mathbf{Q}_n^{-1} + C\mathbf{R}_n^{-1}C'$$

APPENDIX C
PROOF OF THEOREM 2

A. Definitions

Let $\Pi(X) \triangleq \{\rho \in \mathbb{R}^X : \mathbf{1}_X'\rho = 1, 0 \leq \rho(i) \leq 1$ for all $i \in \{1, 2, \ldots, X\}\}$ denote the $X - 1$ dimensional unit simplex.

Definition 2. *First-Order Stochastic Dominance ($\geq_{sd}$): Let $\rho_1, \rho_2 \in \Pi(X)$ be any two belief state vectors. Then $\rho_1 \geq_{sd} \rho_2$ if*

$$\sum_{i=j}^{X}\rho_1(i) \geq \sum_{i=j}^{X}\rho_2(i) \text{ for } j \in \{1, \ldots, X\}.$$

Definition 3. *Second-Order Stochastic Dominance ($\geq_{ssd}$): Let $\rho_1, \rho_2 \in \Pi(X)$ be any two belief state vectors with F_1 and F_2 as the corresponding cumulative distribution functions. Then $\rho_1 \geq_{ssd} \rho_2$ if*

$$\sum_{j=1}^{i}F_1(i) \leq \sum_{j=1}^{i}F_2(i) \text{ for } i \in \{1, \ldots, X\}.$$

Lemma 4. [46] $\rho_2 \geq_s \rho_1$ *iff for all* $v \in \mathcal{V}$, $v'\rho_2 \leq v'\rho_1$, *where* $\mathcal{V}$ *denotes the space of* X- *dimensional vectors* v, *with non-increasing components, i.e,* $v_1 \geq v_2 \geq \ldots v_X$.

Lemma 5. [46] $\rho_2 \geq_s \rho_1$ *iff for all* $v \in \mathcal{V}$, $v'\rho_2 \geq v'\rho_1$, *where* $\mathcal{V}$ *denotes the space of* X- *dimensional vectors* v, *with non-decreasing components, i.e,* $v_1 \leq v_2 \leq \ldots v_X$.

B. Proofs

Theorem 4 ([1])**.** *Any two networks having degree distributions* ρ^1 *and* ρ^2 *respectively, with* $\rho^1 <_{ssd} \rho^2$, *the diffusion threshold* $\lambda^1_* > \lambda^2_*$. [17] $\qquad\square$

In words, as the number of nodes with higher degree increase, the probability of a large fraction of agents becoming infected increases.

Lemma 6. *For any* $p \in [0,1]$, *the transition matrix* $H_k(p)$ *for a preferential attachment graph is such that*

$$H_k^i(p) <_{sd} H_k^{i+1}(p)$$

for $i = 1, 2, \ldots$, *where* $H_k^i(p)$ *denotes the* i^{th} *row of* $H_k(p)$.

Proof of Lemma 6: It is clear from the definition of $H_k(p)$ that each row has only 2 non-zero elements: probability of node undergoing no change and probability of having a degree lesser by 1 and an adjacent edge was added during evolution (25). From the definition of First-order stochastic dominance, we have that a row dominates another row if the tail sum of the former is larger than the latter. Since the matrix $H_k(p)$ is upper bidiagonal, the result follows. $\qquad\square$

Lemma 7. *Let* $H_k(p_1)$ *and* $H_k(p_2)$ *be two social networks modeled using preferential attachment with the probability of adding a new vertex,* $p_i > 0$. *If* $p_1 > p_2$, *then*[18]

$$H_k(p_2) \overset{r}{>}_{sd} H_k(p_1)$$

Proof of Lemma 7: Given $p_1 > 0$, $p_2 > 0$ and $p_1 > p_2$. Since First-order dominance is the comparison of tail sums and since each row has only 2 non-zero elements, to compare the same rows of 2 different matrices, it is sufficient to compare the first element. For the sake of illustration, let us consider row i. Since $i > 0$ and $k > 0$,

$$p_1 > p_2 \Rightarrow 2 - p_1 < 2 - p_2 \Rightarrow i(2-p_1) < i(2-p_2)$$

$$\Rightarrow \frac{i(2-p_1)}{k} < \frac{i(2-p_2)}{k} \Rightarrow 1 - \frac{i(2-p_1)}{k} > 1 - \frac{i(2-p_2)}{k}$$

and therefore $H_k(p_2) \overset{r}{>}_{sd} H_k(p_1)$. $\qquad\square$

Lemma 6 says that the rows of the transition matrix are first-order increasing. Lemma 7 says that the transition probabilities are first-order increasing in the probability of adding new edges, $1 - p$.

---

[17] $<_{ssd}$ denotes second order stochastic dominance.

[18] $\overset{r}{>}_{sd}$ denotes row-wise first order stochastic dominance. $H_k^i(p_2) \;>_{sd}\; H_k^i(p_1)$ for $i = 1, 2, \ldots$.

Lemma 8. *Let* $H_k(p)$ *be such that* $H_k^i(p) <_{sd} H_k^{i+1}(p)$ *for* $i = 1, 2, \ldots$, *where* $H_k^i(p)$ *denotes the* i^{th} *row of* $H_k(p)$. *Then for any probability distributions* ρ_1 *and* ρ_2 *with* $\rho_1 <_{sd} \rho_2$,

$$H_k'(p)\rho_1 <_{sd} H_k'(p)\rho_2$$

Proof of Lemma 8: For convenience, let $\Psi = H_k'(p_2)$. From the definition of First-order dominance on the last row

$$\Psi_{1N^+} \leq \Psi_{2N^+} \leq \Psi_{3N^+} \ldots \Psi_{N^+N^+}$$

From Lemma 5, we have

$$\sum_i \rho_1^i \Psi_{iN^+} \leq \sum_i \rho_2^i \Psi_{iN^+}$$

Any (arbitrary) s^{th} element for the two distribution vectors is given by

$$\sum_i \rho_1^i \sum_{k=s}^{N^+} \Psi_{ik} \quad \text{and} \quad \sum_i \rho_2^i \sum_{k=s}^{N^+} \Psi_{ik}$$

We have from the definition of First-order dominance,

$$\sum_{k=s}^{N^+} \Psi_{ik} \leq \sum_{k=s}^{N^+} \Psi_{(i+1)k} \quad \text{for } i \in \{1, 2 \ldots, N\}$$

From Lemma 5 and using $\rho^1 <_{sd} \rho^2$,

$$\sum_i \rho_1^i \sum_{k=s}^{N^+} \Psi_{ik} \leq \sum_i \rho_2^i \sum_{k=s}^{N^+} \Psi_{ik}$$

$$\sum_{k=s}^{N^+} \sum_i \rho_1^i \Psi_{ik} \leq \sum_{k=s}^{N^+} \sum_i \rho_2^i \Psi_{ik}$$

$$\Psi\rho_1 <_{sd} \Psi\rho_2$$

where ρ_*^i denotes the ith element of the distribution vector ρ_*. $\qquad\square$

Lemma 9. *Let* $p_1 > p_2$ *and* $H_k(p_2) \overset{r}{>}_{sd} H_k(p_1)$. *Then for any probability distribution* ρ,

$$H_k'(p_1)\rho <_{sd} H_k'(p_2)\rho$$

Proof of Lemma 9: For convenience, let $\Psi^2 = H_k'(p_2)$ and $\Psi^1 = H_k'(p_1)$. We know that the maximum degree is represented as N^+. From the definition of First-order dominance on the last row,

$$\Psi_{iN^+}^1 \leq \Psi_{iN^+}^2 \text{ for all } i.$$

$$\Rightarrow \sum_i \rho^i \Psi_{iN^+}^1 \leq \sum_i \rho^i \Psi_{iN^+}^2$$

Let $s \in \{1, 2, \ldots, N\}$ be arbitrary. From the definition of First-order stochastic dominance on corresponding rows of Ψ^1 and Ψ^2,

$$\sum_{k=s} \Psi_{ik}^1 \leq \sum_{k=s} \Psi_{ik}^2 \text{ for all } i.$$

$$\Rightarrow \sum_i \rho^i \sum_{k=s}^{N^+} \Psi_{ik}^1 \leq \sum_i \rho^i \sum_{k=s}^{N^+} \Psi_{ik}^2$$

$$\Rightarrow \sum_{k=s}^{N^+} \sum_i \rho^i \Psi_{ik}^1 \leq \sum_{k=s}^{N^+} \sum_i \rho^i \Psi_{ik}^2$$

$$\Rightarrow \Psi^1 \rho <_{sd} \Psi^2 \rho$$

where ρ^i denotes the ith element of the distribution vector ρ. $\square$

Proof of Theorem 2:

1) The proof of Theorem 2.1 is by induction on k. Consider an initial probability distribution ρ_0.

Base: Let $p_1 > p_2 > 0$. From Lemma 9,

$$\rho_1(p_2) = H_k'(p_2)\rho_0 >_{sd} \rho_1(p_1) = H_k'(p_1)\rho_0$$

By Lemma 8 and Lemma 9,

$$H_k'(p_1)\rho_1(p_1) <_{sd} H_k'(p_1)\rho_1(p_2)$$
$$H_k'(p_1)\rho_1(p_2) <_{sd} H_k'(p_2)\rho_1(p_2)$$
$$\Rightarrow H_k'(p_1)\rho_1(p_1) <_{sd} H_k'(p_2)\rho_1(p_2)$$
$$\Rightarrow (H_k'(p_1))^2\rho_0 <_{sd} (H_k'(p_2))^2\rho_0$$

Induction step: Let the result hold for all $k \le q$.

$$(H_k'(p_2))^q\rho_0 >_{sd} (H_k'(p_1))^q\rho_0$$

Let $(H_k'(p_2))^q\rho_0 = \rho_{q+1}(p_2)$ and $(H_k'(p_1))^q\rho_0 = \rho_{q+1}(p_1)$. By Lemma 8 and Lemma 9,

$$H_k'(p_1)\rho_{q+1}(p_1) <_{sd} H_k'(p_1)\rho_{q+1}(p_2)$$
$$H_k'(p_1)\rho_{q+1}(p_2) <_{sd} H_k'(p_2)\rho_{q+1}(p_2)$$
$$\Rightarrow H_k'(p_1)\rho_{q+1}(p_1) <_{sd} H_k'(p_2)\rho_{q+1}(p_2)$$
$$\Rightarrow (H_k'(p_1))^{q+1}\rho_0 <_{sd} (H_k'(p_2))^{q+1}\rho_0$$

As q is any arbitrary positive integer, the result holds for all $k > 1$. $\square$

2) The proof of Theorem 2.2 easily follows from Theorem 2.1 and Theorem 4 and is omitted. Using Theorem 2, the degree distributions are ordered, first order stochastic dominance implies second order dominance [33], and from Theorem 4, the result follows. $\square$

REFERENCES

[1] D. López-Pintado, "Diffusion in complex social networks," *Games and Economic Behavior*, vol. 62, no. 2, pp. 573–590, 2008.

[2] C. Chamley, *Rational herds: Economic models of social learning*. Cambridge University Press, 2004.

[3] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, pp. 1420–1443, 1978.

[4] E. Mossel and S. Roch, "On the submodularity of influence in social networks," in *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. ACM, 2007, pp. 128–134.

[5] N. Chen, "On the approximability of influence in social networks," *SIAM Journal on Discrete Mathematics*, vol. 23, no. 3, pp. 1400–1415, 2009.

[6] M. O. Jackson, *Social and economic networks*. Princeton university press Princeton, 2008, vol. 3.

[7] D. López-Pintado, "Contagion and coordination in random networks," *International Journal of Game Theory*, vol. 34, no. 3, pp. 371–381, 2006.

[8] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.

[9] F. Vega-Redondo, *Complex social networks*. Cambridge University Press, 2007, no. 44.

[10] J. Zhang and M. J. Moura, "Diffusion in social networks as SIS epidemics: beyond full mixing and complete graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 537–551, 2014.

[11] Z. Wang, W. Zhang, and C. W. Tan, "On inferring rumor source for SIS model under multiple observations," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 2015, pp. 755–759.

[12] V. Krishnamurthy, O. N. Gharehshiran, and M. Hamdi, "Interactive Sensing and Decision Making in Social Networks," *Foundations and Trends in Signal Processing*, vol. 7, no. 1-2, pp. 1–196, 2014.

[13] M. Granovetter, "Network sampling: Some first steps," *American Journal of Sociology*, pp. 1287–1303, 1976.

[14] P. J. Carrington, J. Scott, and S. Wasserman, *Models and methods in social network analysis*. Cambridge university press, 2005, vol. 28.

[15] O. Frank, "Network sampling and model fitting," *Models and methods in social network analysis*, pp. 31–56, 2005.

[16] N. K. Ahmed, J. Neville, and R. Kompella, "Network sampling: From static to streaming graphs," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 8, no. 2, p. 7, 2014.

[17] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[18] D. D. Heckathorn, "Respondent-driven sampling: a new approach to the study of hidden populations," *Social problems*, vol. 44, no. 2, pp. 174–199, 1997.

[19] ——, "Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations," *Social problems*, vol. 49, no. 1, pp. 11–34, 2002.

[20] J. Leskovec and C. Faloutsos, "Sampling from large graphs," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 631–636.

[21] M. Hernández-González and M. V. Basin, "Discrete-time filtering for nonlinear polynomial systems over linear observations," *International Journal of Systems Science*, vol. 45, no. 7, pp. 1461–1472, 2014.

[22] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on signal processing*, vol. 46, no. 5, pp. 1386–1396, 1998.

[23] M. O. Jackson and L. Yariv, "Diffusion of behavior and equilibrium properties in network games," *The American economic review*, vol. 97, no. 2, pp. 92–98, 2007.

[24] M. O. Jackson and B. W. Rogers, "Relating network structure to diffusion properties through stochastic dominance," *The BE Journal of Theoretical Economics*, vol. 7, no. 1, 2007.

[25] M. Hamdi, V. Krishnamurthy, and G. Yin, "Tracking a Markov-modulated stationary degree distribution of a dynamic random graph," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6609–6625, 2014.

[26] F. Chung and L. Lu, *Complex Graphs and Networks*. American Mathematical Society, Providence, 2006, vol. 107.

[27] J. Leskovec, D. Chakrabarti, J. Kleinberg, and C. Faloutsos, "Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication," in *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer, 2005, pp. 133–145.

[28] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[29] A.-L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web," *Physica A: Statistical Mechanics and its Applications*, vol. 281, no. 1, pp. 69–77, 2000.

[30] L. A. Adamic and B. A. Huberman, "Power-law distribution of the world wide web," *Science*, vol. 287, no. 5461, pp. 2115–2115, 2000.

[31] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 591–600.

[32] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan, "Epidemiological modeling of news and rumors on Twitter," in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*. ACM, 2013, p. 8.

[33] V. Krishnamurthy, *Partially Observed Markov Decision Processes*. Cambridge University Press, 2016.

[34] S. Goel and M. J. Salganik, "Respondent-driven sampling as Markov chain Monte carlo," *Statistics in Medicine*, vol. 28, no. 17, pp. 2202–2229, 2009.

[35] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," *Computer networks*, vol. 33, no. 1, pp. 309–320, 2000.

[36] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, pp. 415–444, 2001.

[37] H. Kushner, *Weak convergence methods and singularly perturbed stochastic control and filtering problems*. Springer Science & Business Media, 2012.

[38] J. D. Hamilton, *Time series analysis*. Princeton: Princeton university press, 1994, vol. 2.

[39] Z. Papacharissi and M. de Fatima Oliveira, "Affective news and networked publics: The rhythms of news storytelling on #Egypt," *Journal of Communication*, vol. 62, no. 2, pp. 266–282, 2012.

[40] R. Crane and D. Sornette, "Robust dynamic classes revealed by measuring the response function of a social system," *Proceedings of the National Academy of Sciences*, vol. 105, no. 41, pp. 15 649–15 653, 2008.

[41] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[42] A. Java, X. Song, T. Finin, and B. Tseng, "Why we Twitter: understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 2007, pp. 56–65.

[43] J. Alstott, E. Bullmore, and D. Plenz, "powerlaw: a Python package for analysis of heavy-tailed distributions," *PloS one*, vol. 9, no. 1, p. e85777, 2014.

[44] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[45] M. Benaim and J. Weibull, "Deterministic approximation of stochastic evolution in games," *Econometrica*, vol. 71, no. 3, pp. 873–903, 2003.

[46] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*. Wiley, 2002, vol. 389.