A Polya Urn-Based Model for Epidemics on Networks

Mikhail Hayhoe Fady Alajaji Bahman Gharesifard

***Abstract*— A network epidemics model based on the classical Polya urn scheme is investigated. Temporal contagion processes are generated on the network nodes via a modified Polya sampling scheme that accounts for spatial infection among neighboring nodes. The stochastic properties and the asymptotic behavior of the resulting network contagion process are analyzed. Unlike the classical Polya process, the network process is noted to be non-stationary in general, although it is shown to be time-invariant in its first and some of its second-order statistics and to satisfy martingale convergence properties for certain network structures. Finally, two classical Polya processes, one computational and one analytical, are proposed to statistically approximate the contagion process of each node, showing a good fit for a range of system parameters.**

I. Introduction

We study the dynamics and properties of a contagion process, or *epidemic*, on a network. In this framework an epidemic can represent a disease [1], a computer virus [2], the spread of an innovation, rumor or idea [3], or the dynamics of competing opinions in a social network [4].

Many epidemic models for the study of infection propagation and curing exist in the literature. Our model has similarities with the well-known susceptible-infected-susceptible infection model [5]. In this model, all nodes may initially be healthy or infected. As the epidemic spreads, infected nodes can be cured and become healthy, but healthy nodes may become infected at any time, regardless of whether they have previously been cured. The model that we present is an adaptation of the Polya contagion process [6], [7] to a network setting. Epidemics on networks have been intensively studied in recent years; see [8] and references therein and thereafter. The classical Polya model has been used in a wide range of applications; e.g., see [9] for a summary. In this work, we consider a Polya contagion process for networks by accounting for spatial infection between nodes and examine its stochastic evolution.

We introduce a novel framework for studying epidemics on networks. Our model is motivated by the classical Polya contagion process generated by sampling from an urn containing a finite number of red and black balls [6], [7]. Similar to that setting, each node of the underlying network is equipped with an individual urn; however, instead of drawing solely from its own urn when generating its contagion process, each node has a "super urn", created by combining all the balls in its own urn with the balls in its neighbors' urns. In this sense, the model captures the fact that having infected neighbous increases the chance of an individual node being infected in the future. This concept of the super urn sampling mechanism for incorporating spatial interactions between each node and its neighbors was originally introduced in [10] in the context of the image segmentation and labeling problem. We herein adopt the image model of [10] for a network setting and analyze the resulting contagion process affecting each node of the network for the purpose of epidemic mitigation and control.

More specifically, we study the time evolution and stochastic properties of the proposed network contagion process. We derive an expression for the temporal n-fold joint probability distribution of the process. We show that this process, unlike the classical Polya urn process, is in general non-stationary, and hence not exchangeable. For the special case of complete networks, we analytically find the 1-dimensional and 2-dimensional $(n, 1)$-step marginal distributions of the contagion process. These results show that, even though it is not stationary, the process is nevertheless identically distributed with its later two marginal distributions being invariant to time shifts. We also establish two martingale properties (one for general networks and one for regular networks) regarding the network urn compositions, proving that the proportions of red balls in each node's urn as well as the network average urn proportion converge almost surely to a limit as time grows without bound. These results are useful in studying curing policies. We next provide two approximations to the network contagion process by modeling each node's contagion process via the classical stationary Polya process [6]. In the first one, we approximate each node's process with the classical Polya process whose correlation parameter is empirically selected so that the Kullback-Leibler divergence between its n-fold joint distribution and that of the original node process is minimized. In the second approximation, we propose an analytical classical Polya model whose parameters are chosen by matching its first and $(n, 1)$-step second-order statistics with those of the original node process. Finally, we present ideas for controlling epidemics by studying the average infection in the context of "contagion dilution". We also formulate some control problems with budget constraints for minimizing the limiting average infection. Preliminary simulation results are presented. Throughout, the proofs are omitted for reasons of space and will appear elsewhere.

The authors are with the Department of Mathematics and Statistics at Queen's University, Kingston, ON, Canada m.hayhoe@queensu.ca, fady@mast.queensu.ca, bahman@mast.queensu.ca. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada.

II. Preliminaries

For a sequence $(v_1, ..., v_n)$, we use the notation v^t with $1 \leq t \leq n$ to denote the vector $(v_1, v_2, ..., v_t)$. Our technical results rely on notions from stochastic processes, some of

which we recall here. Throughout, we assume that the reader is familiar with basic notions of probability theory.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and consider the stochastic process $\{Z_n\}_{n=1}^{\infty}$, where Z_n is a random variable on Ω. We often refer to the indices of the process as "time" indices. We recall that the process $\{Z_n\}_{n=1}^{\infty}$ is *stationary* if for any $n \in \mathbb{Z}_{\geq 1}$, its n-fold joint probability distribution (i.e., the distribution of $(Z_1, ..., Z_n)$) is invariant to time shifts. Further, $\{Z_n\}_{n=1}^{\infty}$ is *exchangeable* if for any $n \in \mathbb{Z}_{\geq 1}$, its n-fold joint distribution is invariant to permutations of the indices $1, ..., n$. It directly follows from the definitions that an exchangeable process is stationary. Lastly, the process $\{Z_n\}_{n=1}^{\infty}$ is called a *martingale* with respect to $\{X_n\}_{n=1}^{\infty}$ if $E[|Z_n|] < \infty$ and $E[Z_{n+1}|X^n] = Z_n$ almost surely, for all n. If $E[Z_{n+1}|Z^n] = Z_n$ almost surely, we say that $\{Z_n\}_{n=1}^{\infty}$ is a martingale with respect to itself and simply state that it is a martingale. Precise definitions of all notions, including that of *ergodicity* can be found in standard texts (e.g., [11]).

We now recall the classical version of the Polya contagion process [6], [7]. Consider an urn with $R \in \mathbb{Z}_{>0}$ red balls and $B \in \mathbb{Z}_{>0}$ black balls. We denote the total number of balls by T, i.e., $T = R + B$. At each time step, a ball is drawn from the urn. The ball is then returned along with $\Delta > 0$ balls of the same color. We use an indicator Z_n to denote the color of ball in the nth draw: $Z_n = 1$ if the nth draw is red, and $Z_n = 0$ if the nth draw is black. Let U_n denote the proportion of red balls in the urn after the nth draw. Then

$$U_n := \frac{\rho_c + (Z_n + Z_{n-1} + \cdots + Z_1)\delta_c}{1 + n\delta_c}$$

where $\rho_c = \frac{R}{T}$ is the initial proportion of red balls in the urn and $\delta_c = \frac{\Delta}{T}$ is a correlation parameter. The above classical Polya process $\{Z_n\}_{n=1}^{\infty}$ is fully described by its parameters ρ_c and δ_c; we denote it by $\mathsf{Polya}(\rho_c, \delta_c)$. The conditional probability of drawing a red ball at time n, given $Z^{n-1} = (Z_1, \cdots, Z_{n-1})$, is given by $P(Z_n = 1 \mid Z^{n-1}) = U_{n-1}$.

It can be easily shown that $\{U_n\}_{n=1}^{\infty}$ is a martingale [12]. The process $\{Z_n\}_{n=1}^{\infty}$, whose n-fold joint distribution is denoted by $Q^{(n)}_{\rho_c,\delta_c}$, is also exchangeable (hence stationary) and non-ergodic with both U_n and the process sample average $\frac{1}{n}\sum_{i=1}^{n} Z_i$ converging almost surely as $n \to \infty$ to a random variable governed by the Beta distribution with parameters $\frac{\rho_c}{\delta_c}$ and $\frac{1-\rho_c}{\delta_c}$; we denote this distribution by $\mathsf{Beta}(\frac{\rho_c}{\delta_c}, \frac{1-\rho_c}{\delta_c})$ [12], [13]. Lastly, the 1-dimensional distribution of the Polya process is $Q^{(1)}_{\rho_c,\delta_c}(a) = P(Z_n = a) = (\rho_c)^a(1-\rho_c)^{1-a}$, for all $n \in \mathbb{Z}_{\geq 1}$ and $a \in \{0, 1\}$.

III. Network Contagion Process

In this section, we introduce a generalization of the Polya contagion process to scenarios on graphs, where each individual node in the graph is still equipped with an urn; however, the node's neighboring structure affects the evolution of its process. This model hence captures spatial contagion, where infected neighbors increase the chance of a node being infected in the future. Consider an undirected graph $\mathcal{G} = (V, \mathcal{E})$, where $V = \{1, \ldots, N\}$ is the set of $N \in \mathbb{Z}_{\geq 1}$ nodes and $\mathcal{E} \subset V \times V$. We assume that $\mathcal{G}$ is connected, i.e., there is a path between any two nodes in $\mathcal{G}$. We use $\mathcal{N}_i$ to denote the set of nodes that are neighbors to node i, that is $\mathcal{N}_i = \{v \in V : (i, v) \in \mathcal{E}\}$, and $\mathcal{N}_i' = \{i\} \cup \mathcal{N}_i$. Each node $i \in V$ is equipped with an urn, initially with $R_i \in \mathbb{Z}_{>0}$ red balls and $B_i \in \mathbb{Z}_{>0}$ black balls (we do not let $R_i = 0$ or $B_i = 0$ to avoid any degenerate cases). We let $T_i = R_i + B_i$ be the total number of balls in the ith urn, $i = 1, \cdots, N$. We use $Z_{i,n}$ as an indicator for the ball drawn for node i at time n: $Z_{i,n} = 1$ if the nth draw for node i is red, and $Z_{i,n} = 0$ if the nth draw for node i is black. However, instead of drawing solely from its own urn, each node has a "super urn" created by combining all the balls in its own urn with the balls in its neighbors' urns; see Fig. 1. This allows the spatial relationships between nodes to influence their state of infection. This means that $Z_{i,n}$ is the indicator for a ball drawn from node i's super urn, and not its individual urn. Hence, the super urn of node i has $\bar{R}_i = \sum_{j \in \mathcal{N}_i'} R_j$ red balls, $\bar{B}_i = \sum_{j \in \mathcal{N}_i'} B_j$ black balls, and $\bar{T}_i = \sum_{j \in \mathcal{N}_i'} T_j$ balls in total.

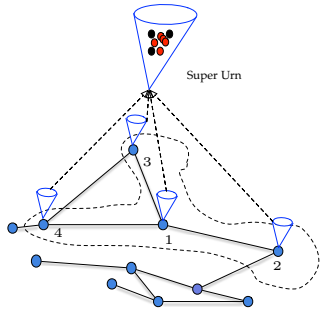


Fig. 1. Illustration of a super urn in a network, here for node 1.

We further consider a time-varying version of the classical Polya contagion process, following [14], where at time t for node $i \in V$, $\Delta_{r,i}(t)$ net red balls are added to node i's urn when a red ball is drawn, and $\Delta_{b,i}(t)$ net black balls are added to node i's urn when a black ball is drawn. When $\Delta_{r,i}(t) = \Delta_{b,i}(t)$ for all t, we write $\Delta_i(t)$ instead; if the Δ's are not node-dependent, we omit the node index. We assume throughout that $\Delta_{r,i}(t) \geq 0, \Delta_{b,i}(t) \geq 0$, for all t and that there exists $i \in V$ and t such that $\Delta_{r,i}(t) + \Delta_{b,i}(t) \neq 0$; otherwise we are simply sampling with replacement. To express the proportion of red balls in the individual urns of the nodes, we define the random vector $U_n = (U_{1,n}, \ldots, U_{N,n})$ after the nth draw, where $U_{i,n}$ is the proportion of red balls in node i's urn after the nth draw, $i \in V$. For node i,

$$U_{i,n} := \frac{R_i + \sum_{t=1}^{n} Z_{i,t}\Delta_{r,i}(t)}{T_i + \sum_{t=1}^{n} Z_{i,t}\Delta_{r,i}(t) + (1 - Z_{i,t})\Delta_{b,i}(t)}$$

where the numerator represents the total number of red balls in node i's urn after the nth draw, while the denominator is the total number of balls in the same urn. In the context of epidemics, the red and black balls in an urn, respectively, represent "infection" and "healthiness". Let now

$$X_{j,n} = T_j + \sum_{t=1}^{n} Z_{j,t}\Delta_{r,j}(t) + (1 - Z_{j,t})\Delta_{b,j}(t).$$

Then the conditional probability of drawing a red ball from the super urn of node i at time n given the complete network history, i.e. given all the past $n-1$ draw variables for each node in the network $\{Z_j^{n-1}\}_{j=1}^N = \{(Z_{1,1},\cdots,Z_{1,n-1}),\cdots,(Z_{N,1},\cdots,Z_{N,n-1})\}$, satisfies

$$P\left(Z_{i,n}=1|\{Z_j^{n-1}\}_{j=1}^N\right) = \frac{\bar{R}_i + \sum_{j\in\mathcal{N}_i'}\sum_{t=1}^{n-1} Z_{j,t}\Delta_{r,j}(t)}{\sum_{j\in\mathcal{N}_i'} X_{j,n-1}}$$
$$= \frac{\sum_{j\in\mathcal{N}_i'} U_{j,n-1}X_{j,n-1}}{\sum_{j\in\mathcal{N}_i'} X_{j,n-1}}$$
$$=: g_{i,n}(Z_1^{n-1},\cdots,Z_N^{n-1}). \quad (1)$$

A main objective throughout the rest of this paper is to study the evolution and stochastic properties of the process defined above. Using the above conditional probability, we can determine the n-fold joint probability of the entire network $\mathcal{G}$: for $a_i^n \in \{0,1\}^n$, $i=1,\cdots,N$, we have that

$$P_{\mathcal{G}}^{(n)}(a_1^n,\cdots,a_N^n) := P\left(\{Z_i^n = a_i^n\}_{i=1}^N\right)$$
$$= \prod_{t=1}^n P\left(Z_{1,t}=a_{1,t},...,Z_{N,t}=a_{N,t} \mid \{Z^{t-1}=a^{t-1}\}_{i=1}^N\right)$$
$$= \prod_{t=1}^n\prod_{i=1}^N (g_{i,t})^{a_{i,t}}(1-g_{i,t})^{1-a_{i,t}}, \quad (2)$$

where $g_{i,t}$ is defined in (1). Similar to the classical Polya urn process, we are interested in studying the asymptotic behavior of each node's contagion process, since understanding many interesting questions regarding the limiting behavior of epidemics on networks and formulating curing strategies are closely related to this problem. With the above explicit joint distribution, it is possible to determine the distributions of each node's process. More specifically, using (2), the n-fold distribution of node i's process at time $t \geq n$ is

$$P_{i,t}^{(n)}(a_{i,t-n+1},...,a_{i,t}) := \sum_{\substack{a_i^{t-n}\in\{0,1\}^{t-n}\\ a_j^t\in\{0,1\}^t, j\neq i}} P_{\mathcal{G}}^{(n)}(a_1^t,\cdots,a_N^t).$$

We define the *average infection* in the network at time n as

$$\tilde{I}_n := \frac{1}{N}\sum_{i=1}^N P(Z_{i,n}=1) = \frac{1}{N}\sum_{i=1}^N P_{i,n}^{(1)}(1).$$

Unfortunately for an arbitrary network, the above quantity does not yield an exact analytical formula (except in the simple case of complete networks). As such, it is in general hard to mathematically analyze the asymptotic behavior of $\tilde{I}_n$, which we wish to minimize when attempting to cure an epidemic. Instead we examine the asymptotic stochastic behavior of a closely related variable given by the average individual proportion of red balls at time n, $\tilde{U}_n := \frac{1}{N}\sum_{i=1}^N U_{i,n}$, which we call the *network susceptibility*. Indeed in the next section, we derive martingale results for both $\{U_{i,n}\}_{n=1}^\infty$ and $\{\tilde{U}_n\}_{n=1}^\infty$ under certain network configurations.

Remark 3.1: **(Finite Memory):** It is worth pointing out that a practical adaptation to our model can be considered, where urns have "finite memory" in the sense that the balls added after each draw are only kept in each node's urn for a finite number of future draws. This model is developed in [13] for the classical Polya process in the context of modeling communication channels, where it is shown that the resulting finite memory contagion process is stationary, Markovian and ergodic. We leave investigating this scenario in our context to a future work. •

IV. Stochastic Properties

We next examine the stochastic properties of the network contagion process. We assume throughout this section that $\Delta_{r,i}(t) = \Delta_{b,i}(t) = \Delta > 0$, for all $i \in V$ and times t; that is the net number of red and black balls added are equal and constant in time for all nodes. In the case of a complete network, the composition of every nodes' super urn is identical; in a sense, there is only one super urn that is being drawn from. Thus for a complete network the super urn model is analogous to one urn where multiple draws occur with replacement, which has been recently studied in detail [15]. However, the analysis in [15] is carried in an aggregate sense and, in particular, only for the entire urn, not individual processes. Unfortunately, this aggregate approach does not work in a network setting, and so in that case the super urn model is more useful.

We will now derive some stochastic distributions for the complete network; later on, we will derive martingale results for more general networks. Given that the network is complete, we focus on one of the nodes, say $i \in V$. Defining the event $A_{n-1} = \{Z_{i,n-1}=a_{n-1},...,Z_{i,1}=a_1\}$, we can write, using (1) under the above assumptions and omitting some details for reasons of space, that

$$P(Z_{i,n}=1, A_{n-1})$$
$$= \sum_{b_j^{n-1}\in\{0,1\}^{n-1}, j\neq i} P(Z_{i,n}=1|A_{n-1},\{Z_{j,1}^{n-1}=b_j^{n-1}\}_{j\neq i})$$
$$\times P(A_{n-1},\{Z_j^{n-1}=b_j^{n-1}\}_{j\neq i})$$
$$= \frac{\sum_{t=1}^{n-1}\left[a_t P(A_{n-1}) + \sum_{j\neq i} P(A_{n-1},Z_{j,t}=1)\right]\frac{\delta}{N}}{1+(n-1)\delta}$$
$$+ \frac{\rho P(A_{n-1})}{1+(n-1)\delta} \quad (3)$$

where $\rho = \frac{\sum_{i=1}^N R_i}{\sum_{i=1}^N T_i}$ and $\delta = \frac{N\Delta}{\sum_{i=1}^N T_i}$. Then, by summing out (3) over $a^{n-1} \in \{0,1\}^{n-1}$ we obtain

$$P(Z_{i,n}=1) = \frac{\rho + \frac{\delta}{N}\sum_{t=1}^{n-1}\sum_{j=1}^N P(Z_{j,t}=1)}{1+(n-1)\delta}. \quad (4)$$

An interesting corollary of this derivation is stated next.

Lemma 4.1: **(Complete Network Marginal Distribution):** The 1-dimensional marginal distribution of node i's contagion draw process $\{Z_{i,n}\}_{n=1}^\infty$ for the N-node complete network is given by

$$P_{i,n}^{(1)}(a) = P(Z_{i,n}=a) = \rho^a(1-\rho)^{1-a}.$$

where $i \in V$, $n \geq 1$, and $a \in \{0,1\}$.

We next show that process is not stationary in general.

Remark 4.2: **(Non-Stationarity of the Network Contagion Process):** Consider a 2-node complete network. Then, using (2), one can obtain that

$$P(Z_{1,2}=1,Z_{1,1}=1)=\rho\frac{\rho+(1+\rho)\frac{\delta}{2}}{1+\delta},$$
$$P(Z_{1,3}=1,Z_{1,2}=1)=\rho\frac{4\rho+\delta(2+14\rho)+\delta^2(6+14\rho)}{4(1+\delta)^2(1+2\delta)}+\rho\frac{\delta^3(5+3\rho)}{4(1+\delta)^2(1+2\delta)}.$$

and hence the network process is not stationary. •

Since every exchangeable process is stationary, Remark 4.2 implies that the process is also not exchangeable in general. However, there are still some notions of stationarity present; for example, our next result show how the relationship between the first and nth draws in a complete network is consistent in time.

Lemma 4.3: **(Complete Network $(n,1)$-step Marginal Distribution):** For the complete network, the 2-dimensional marginal distribution that node i's draw variables at times n and 1 are both one is given by

$$P(Z_{i,n}=1,Z_{i,1}=1)=\rho\frac{\rho+(1+(N-1)\rho)\frac{\delta}{N}}{1+\delta}$$

for $i\in V$, $n\geq 2$. Furthermore, for any other node k,

$$P(Z_{k,n}=1,Z_{i,1}=1)=\rho\frac{\rho+(1+(N-1)\rho)\frac{\delta}{N}}{1+\delta}.$$

Although the process is not stationary in general, our simulated results suggest that it is in fact "asymptotically stationary", in the sense that after some sufficient "settling" time, the joint probabilities are invariant to shifts in the time indices. A representative example is shown in Fig. 2.

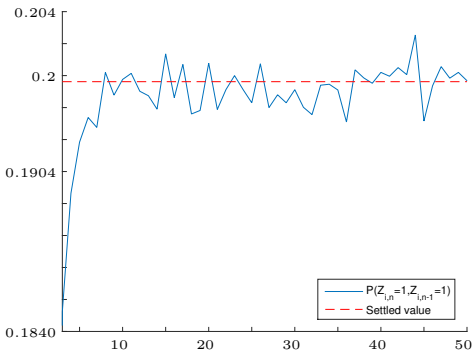


Fig. 2. Simulated values for $P(Z_{1,n}=1,Z_{1,n-1}=1)$ for $2\leq n\leq 50$ averaged over 50,000 simulated trials, each ran until $n=1000$. This is on a 2-node network with $R_1=R_2=4, B_1=B_2=8, \Delta=47$, and the settled value is about 0.1996. Here we observe asymptotic stationarity, as for some large enough n the deviations from the settled value are very small in magnitude.

We now turn our attention to the martingale properties of the network contagion process, where we do not assume that the network is necessarily complete.

Theorem 4.4: **(General Network Individual Urn Proportion Martingale):** For a general network $\mathcal{G}=(V,\mathcal{E})$, $\Delta_{r,i}(n)=\Delta_{b,i}(n)=\Delta$ and $T_i=T$, for all $i\in V$ and all n, the individual proportion of red balls $\{U_{i,n}\}_{n=1}^{\infty}$ is a martingale with respect to $\{U_n\}_{n=1}^{\infty}$, where $U_n=(U_{1,n},\cdots,U_{N,n})$ if and only if

$$\frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}U_{j,n}=U_{i,n}$$

for all n, almost surely.

If the condition in Theorem 4.4 holds, we obtain by the martingale convergence theorem [11], that for any i, both $U_{i,n}$ and $\frac{1}{n}\sum_{t=1}^{n}Z_{i,t}$ converge almost surely to a limit as $n\to\infty$. However, the condition of Theorem 4.4, barring the trivial single node scenario (which reverts to the classical Polya scheme), is not verifiable. In fact, it is not clear that this ever occurs in general. This bodes well for investigations into curing, since if the condition in Theorem 4.4 held then we would have no hope of curing the infection beyond the initial level ρ. To achieve a more applicable result, we now examine the evolution of the network susceptibility $\tilde{U}_n=\frac{1}{N}\sum_{i=1}^{N}U_{i,n}$ when the network is regular.

Theorem 4.5: **(Regular Network Susceptibility Martingale):** For a regular network $\mathcal{G}=(V,\mathcal{E})$, i.e., $|\mathcal{N}_i|=|\mathcal{N}_j|$ for all $i,j\in V$, with $\Delta_{r,i}(n)=\Delta_{b,i}(n)=\Delta$ and $T_i=T$ for all nodes $i\in V$ and times n, the average urn proportion of red balls $\{\tilde{U}_n\}_{n=1}^{\infty}$, where $\tilde{U}_n=\frac{1}{N}\sum_{i=1}^{N}U_{i,n}$, is a martingale with respect to $\{U_n\}_{n=1}^{\infty}$, and so there exists a random variable M such that $\{\tilde{U}_n\}$ converges almost surely to M as $n\to\infty$.

V. Model Approximations

As previously noted, the dynamics of the network contagion process are complicated, especially when considered on general networks. For this reason, in this section we develop two useful approximations to this process on a general network that allow us to shed some light on its asymptotic behavior. Throughout this section, unless stated otherwise, we consider general network topologies. However, to match the $1-$ and $(n,1)$-step distributions in the models below, we assume that the neighborhood of each node i is a complete sub-network in order to apply Lemmas 4.1 and 4.3.

A. Approximation: Computational Model

We now introduce our first approximation technique, where we approximate the contagion process of each node in the network with a classical Polya urn process.

Model 5.1: **(Computational Model):** Given any node i in the network, we approximate the dynamics of its contagion process using a classical Polya$(\rho_c=\rho,\delta_c=\hat{\delta})$ process with distribution $Q^{(n)}_{\rho_c,\delta_c}=Q^{(n)}_{\rho,\hat{\delta}}$. Here ρ_c is chosen to be the proportion of red balls ρ in the node's super urn, so that the 1-dimensional distributions of the classical Polya process and the node process $\{Z_{i,n}\}$ coincide, while $\hat{\delta}$ is set by performing a minimization to find the value that best fits $Q^{(n)}_{\rho,\hat{\delta}}$ to the distribution of $\{Z_{i,n}\}_{n=1}^{\infty}$ of node $i\in V$. We use a divergence measure, denoted by $D(\cdot||\cdot)$, to observe the quality of the fit. More specifically,

$$\rho=\frac{\sum_{j\in\mathcal{N}_i'}R_j}{\sum_{j\in\mathcal{N}_i'}T_j},\qquad\text{and}\qquad\hat{\delta}=\arg\min_{\delta'}\frac{1}{n}D\left(P^{(n)}_{i,n}||Q^{(n)}_{\rho,\hat{\delta}}\right)$$

where

$$Q^{(n)}_{\rho,\tilde{\delta}}(a_1,...,a_n) = \frac{\Gamma\left(\frac{1}{\tilde{\delta}}\right)\Gamma\left(\frac{\rho}{\tilde{\delta}}+\bar{a}^n\right)\Gamma\left(\frac{1-\rho}{\tilde{\delta}}+n-\bar{a}^n\right)}{\Gamma\left(\frac{1}{\tilde{\delta}}+n\right)\Gamma\left(\frac{\rho}{\tilde{\delta}}\right)\Gamma\left(\frac{1-\rho}{\tilde{\delta}}\right)}$$

where $\Gamma(\cdot)$ is the Gamma function and $\bar{a}^n = a_1 + \cdots + a_n$.

The explicit derivation of the distribution $Q^{(n)}_{\rho,\tilde{\delta}}$ can be found in [12], [16]. This method ensures that the fit of $Q^{(n)}_{\rho,\tilde{\delta}}$ is as close as possible under the given divergence measure. To simplify the calculation, we use the Kullback-Leibler divergence [17]; we thus have that

$$\begin{aligned}\hat{\delta} &= \arg\min_{\tilde{\delta}} \frac{1}{n} \sum_{a^n\in\{0,1\}^n} P^{(n)}_{i,n}(a^n) \log\left(\frac{P^{(n)}_{i,n}(a^n)}{Q^{(n)}_{\rho,\tilde{\delta}}(a^n)}\right) \\ &= \arg\max_{\tilde{\delta}} \frac{1}{n} \sum_{a^n\in\{0,1\}^n} P^{(n)}_{i,n}(a^n) \log Q^{(n)}_{\rho,\tilde{\delta}}(a^n)\end{aligned}$$

since $P^{(n)}_{i,n}(a^n)\log P^{(n)}_{i,n}(a^n)$ is independent of $\tilde{\delta}$. The approximating process is stationary and exchangeable since it is a classical Polya process. We also know (from Section II) that it is non-ergodic with its sample average converging almost surely to the $\mathsf{Beta}(\frac{\rho}{\delta}, \frac{1-\rho}{\delta})$ distribution. Calculating an analytic expression for the minimizing $\hat{\delta}$ is not feasible in general, and hence should be performed computationally. However, due to the above minimization, the value of $\hat{\delta}$ is, by definition, the best way to fit a Polya process to the process $\{Z_{i,n}\}_{n=1}^{\infty}$ for a given n.

B. Approximation: Analytical Model

An alternative to Model 5.1 is to attempt to find an approximation whose parameters can be determined analytically.

Model 5.2: **(Analytical Model):** For any given node i, we approximate the dynamics of its process $\{Z_{i,n}\}_{n=1}^{\infty}$ by using a classical $\mathsf{Polya}(\rho_c = \rho, \delta_c = \delta')$ process, where

$$\rho = \frac{\sum_{j\in\mathcal{N}_i'} R_j}{\sum_{j\in\mathcal{N}_i'} T_j}, \quad \text{and} \quad \delta' = \frac{\delta}{N+(N-1)\delta},$$

where $\delta = \frac{N\Delta}{\sum_{j\in\mathcal{N}_i'} T_j}$. Here the parameters of the classical Polya process are chosen by directly matching its first and $(n,1)$-step second-order statistics with those of $\{Z_{i,n}\}_{n=1}^{\infty}$. This method avoids the computational burden of the previous model by yielding an analytical expression for the correlation parameter of the classical Polya process.

We next prove that under some stationarity and symmetry assumptions, the contagion process for a complete network running on each node is statistically identical to the classical Polya process of Model 5.2.

Lemma 5.3: **(Exact Representation):** Suppose that

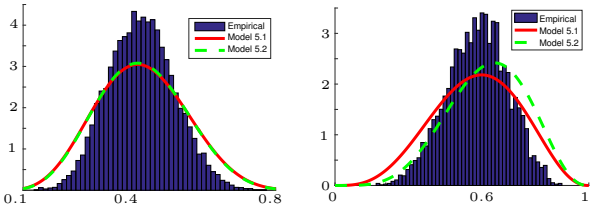
(i) $P(Z_{i,1} = 1 \mid Z_j^n = a^n) = \rho$, and

(ii) $P(Z_{i,t} = 1 \mid Z_j^n = a^n) = P(Z_{k,n+1} = 1 \mid Z_j^n = a^n)$,

for all $n \geq 1, 2 \leq t \leq n$, $i,j,k \in V$, $a^n \in \{0,1\}^n$. Then for any node i in a complete network, $\{Z_{i,n}\}_{n=1}^{\infty}$ is given exactly by the $\mathsf{Polya}(\rho, \delta')$ process.

Unfortunately, in a general network setting neither of these assumptions hold true. However, this result motivates the fact that this analytical approximation is reasonable to use for situations where conditions (i) and (ii) hold within tolerable margins of error; empirical evidence indicates that this occurs for large values of N, since as N increases the quality of the fit improves. This approximation, nevertheless, drastically reduces the complexity in analyzing the individual contagion draw processes, as closed-form expressions for the process parameters are available.

We close this section with numerical demonstrations on the fitness of both approximation models. Fig. 3 shows representative comparisons between the $\mathsf{Beta}(\frac{\rho}{\delta}, \frac{1-\rho}{\delta})$ and $\mathsf{Beta}(\frac{\rho}{\delta'}, \frac{1-\rho}{\delta'})$ pdfs with the simulated histogram of $\frac{1}{n}\sum_{t=1}^{n} Z_{i,n}$, where $n = 1000$, for an arbitrary node i in a network. Recall that the Beta pdfs are the distribution of the limit random variables to which the sample average of the process of Models 5.1 and 5.2 converge almost surely, respectively, as $n \to \infty$ (see Section II). We observe in Fig. 3, for networks where R_i and B_i differ between nodes, that the shape of the pdfs approximately fit the histogram.



(a) Complete network, $N = 2$, $\sum_{i=1}^{N} R_i = 6$, $\sum_{i=1}^{N} B_i = 8$, $\Delta = 1$

(b) Barabasi-Albert network [18], $N = 5$, node degrees are $2, 3, 1, 1, 1$ resp., $\sum_{i=1}^{N} R_i = 28$, $\sum_{i=1}^{N} B_i = 21$, $\Delta = 8$

Fig. 3. Comparison of simulated normalized histogram for $\frac{1}{n}\sum_{t=1}^{n} Z_{i,n}$ for node 1 in a given network (ran until time $1,000$ and averaged over $5,000$ simulated trials), the $\mathsf{Beta}(\frac{\rho}{\delta}, \frac{1-\rho}{\delta})$ pdf from Model 5.1, and the $\mathsf{Beta}(\frac{\rho}{\delta'}, \frac{1-\rho}{\delta'})$ pdf from Model 5.2. The pdfs fit similarly for the complete network, but for the general non-complete network Model 5.1 fits best.

VI. Future Directions: Controlling Epidemics

With a model in place, the next question is to determine strategies to control and mitigate epidemics. Our objective is to study the average infection $\tilde{I}_n$ as n grows without bound; but when seeking analytic results, it might be more amenable to observe the asymptotic behavior of the network susceptibility $\tilde{U}_n$ (note that when $\tilde{U}_n$ decreases, so does $\tilde{I}_n$). In this framework, we wish to reduce the level of infection to some acceptable threshold. We say the epidemic has been *ϵ-eliminated* if $\limsup_{n\to\infty} \tilde{I}_n \leq \epsilon$, $\epsilon \in [0,1]$.

A. Contagion Dilution

In situations where the conditions of Lemma 5.3 hold within an acceptable range of error, making Model 5.2 a good approximation for the nodes' contagion processes, we note that as the number of nodes N increases, the correlation parameter δ' decreases. Indeed, $\delta' \to 0$ as $N \to \infty$, and thus the draw variables of the process of each node become independent and identically distributed, since we are simply drawing with replacement. Hence for each node i, by the strong law of large numbers, we know that the sample

average $\frac{1}{n}\sum_{t=1}^{n} Z_{i,n}$ converges almost surely to a constant, which must be the expected value $E[Z_{i,n}] = \rho$. This means that for complete networks with a large enough number of nodes, the sample average of draws is effectively constant at ρ, and so the average infection is stable and fixed at ρ. This implies that by increasing the number of nodes in the network and by making the network fully connected so that the conditions of Lemma 5.3 hold, we may limit the spread of contagion beyond the initial level of infection ρ. The reduction of contagion spread effectively means that all nodes average out their own individual initial infection and share it in the network. For example, a large and highly connected group of healthy nodes and one very infected node will eventually become a group of slightly less healthy individuals, but none will be very infected. This means that the epidemic will be almost surely ρ-eliminated, but there are no guarantees that it do better, regardless of the initial conditions. One can interpret the outcome of this discussion in the framework of consensus or opinion dynamics, where contagion dilution drastically reduces the opinion of outliers with extreme views.

B. Establishing a Control Problem

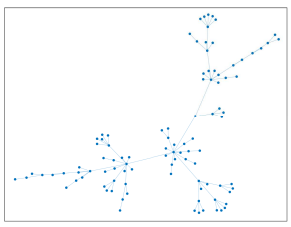
We note that the $\Delta_{b,i}(\cdot)$ quantities, which denote the net number of "healthy" balls added to node i's urn after each draw (see Section III), can play the role of "healing or curing parameters". When these parameters are judiciously selected subject to an allowable budget, say $\mathcal{B}$, on the maximal number of healthy balls that can be added in the network at any given time, the network epidemic can be steered towards a desirable level. For a given ϵ, healing budget $\mathcal{B}$ and network $\mathcal{G}$, we say that curing from initial conditions $\{R_i, B_i\}_{i\in V}$ is *ϵ-achievable* if there exists a $\mathcal{B}$-constrained choice of $\{\Delta_{b,i}(\cdot)\}_{i\in V}$ such that the epidemic can be ϵ-eliminated on $\mathcal{G}$. We thus consider the following problem.

Problem 6.1: **(Average Infection Budget Constraint):** Minimize the limiting average infection $\tilde{I}_n$ subject to a budget $\mathcal{B}$ on the total healing at each time step:

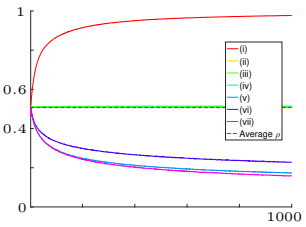
$$\min_{\substack{\sum_{i=1}^{N}\Delta_{b,i}(t)\le\mathcal{B}\\ \forall t}} \limsup_{n\to\infty} \tilde{I}_n$$

In the case of regular graphs under some conditions, using Theorem 4.5, we know that the limit of the related measure $\tilde{U}_n$, the network susceptibility, exists. In the general case, we simply attempt to reduce infection in the worst-case scenario. The solution to Problem 6.1 would be an infinite horizon optimal control policy that would yield the best possible level of epidemic elimination for a given network, budget, and initial conditions. However, finding such a policy in general appears to be difficult. Although discussion of potential policies or numerical approximations are left to a future work, we finish this paper by comparing a number of potential strategies in Fig. 4 to show how the choice of allocation of curing can affect the outcome. The figure indicates strategies which incorporate the degree of the nodes perform best. Strategies used for the choice of $\Delta_{b,i}(\cdot)$, in increasing order of empirical effectiveness, include:

(i) Curing only the most infected node: $\Delta_{b,i}(t) = \mathcal{B}$ if $\arg\max_{i\in V} U_{i,t-1} = i$ and 0 otherwise.
(ii) Ratio of infectedness: $\Delta_{b,i}(t) = \mathcal{B}\frac{U_{i,t-1}}{\sum_{i=1}^{N} U_{j,t-1}}$
(iii) Super urn ratio: $\Delta_{b,i}(t) = \mathcal{B}\frac{\bar{R}_i/\bar{T}_i}{\sum_{j=1}^{N}\bar{R}_j/\bar{T}_j}$
(iv) Uniformly applying the budget: $\Delta_{b,i}(t) = \frac{\mathcal{B}}{N}$
(v) Node degree: $\Delta_{b,i}(t) = \mathcal{B}\frac{|\mathcal{N}_i|}{\sum_{i=1}^{N}|\mathcal{N}_i|}$
(vi) Degree and infectedness: $\Delta_{b,i}(t) = \mathcal{B}\frac{|\mathcal{N}_i|U_{i,t-1}}{\sum_{i=1}^{N}|\mathcal{N}_i|U_{i,t-1}}$
(vii) Degree and super urn ratio: $\Delta_{b,i}(t) = \mathcal{B}\frac{|\mathcal{N}_i|\bar{R}_i/\bar{T}_i}{\sum_{j=1}^{N}|\mathcal{N}_j|\bar{R}_j/\bar{T}_j}$.



(a) Barabasi-Albert network [18] used for simulations, $N = 100$



(b) Comparison of curing strategies (lower values mean less infection)

Fig. 4. Comparison of the average infection $\tilde{I}_n$, $1 \le n \le 1,000$, for different curing strategies under identical initial conditions on a 100-node Barabasi-Albert network. Here the initial conditions are $\sum_{i=1}^{N} R_i = 587, \sum_{i=1}^{N} B_i = 568$, and so $\rho = 0.5082$. All nodes use $\Delta_r = 4$, the budget is $\mathcal{B} = 400$, and results are averaged over $2,000$ trials. Strategies (ii)-(iv) and (vi)-(vii) perform nearly identically, and are overlaid.

REFERENCES

[1] L. Kim, M. Abramson, K. Drakopoulos, S. Kolitz, and A. Ozdaglar, "Estimating social network structure and propagation dynamics for an infectious disease," in *Proc. Int. Conf. Social Computing, Behavioral-Cultural Modeling, and Prediction*, pp. 85–93, Springer, 2014.
[2] M. Garetto, W. Gong, and D. Towsley, "Modeling malware spreading dynamics," in *Proc. IEEE Int. Conf. Comp. Commun.*, vol. 3, pp. 1869–1879, 2003.
[3] E. M. Rogers, *Diffusion of Innovations*. Simon and Schuster, 2003.
[4] E. Adar and L. A. Adamic, "Tracking information epidemics in blogspace," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intelligence*, pp. 207–214, 2005.
[5] D. Easley and J. Kleinberg, *Networks, Crowds and Markets: Reasoning about a Highly Connected World*. Cambridge Univ. Press, 2010.
[6] G. Polya, "Sur quelques points de la théorie des probabilités," *Annales de l'institut Henri Poincaré*, vol. 1, no. 2, pp. 117–161, 1930.
[7] F. Eggenberger and G. Polya, "Über die statistik verketteter vorgänge," *Z. Angew. Math. Mech.*, vol. 3, no. 4, pp. 279–289, 1923.
[8] P. V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 1–14, 2009.
[9] R. Pemantle, "A survey of random processes with reinforcement," *Probab. Surveys*, vol. 4, no. 0, pp. 1–79, 2007.
[10] A. Banerjee, P. Burlina, and F. Alajaji, "Image segmentation and labeling using the Polya urn model," *IEEE Trans. Image Proc.*, vol. 8, no. 9, pp. 1243–1253, 1999.
[11] R. Ash and C. Doléans-Dade, *Probability and Measure Theory*. Academic Press, 2000.
[12] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2. New York: Wiley, 2 ed., 1971.
[13] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2035–2041, 1994.
[14] R. Pemantle, "A time-dependent version of Polya's urn," *Journal of Theor. Probab.*, vol. 3, no. 4, pp. 627–637, 1990.
[15] M. Chen and M. Kuba, "On generalized Polya urn models," *Journal of Appl. Probab.*, vol. 50, no. 4, pp. 1169–1186, 2013.
[16] N. L. Johnson and S. Kotz, *Urn Models and Their Application*. 1977.
[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2006.
[18] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, 2002.