Network Modeling of Consumers' Selection of Providers Based on Online Reviews

Tian Gan, Rishita Das, and Maurizio Porfiri, *Fellow, IEEE*

***Abstract*—Information spreading over online review systems affects people's opinions and choices. In this work, we study consumers' decision-making process with respect to provider selection, accounting for the providers' online reviews and accessibility. We propose a network-based dynamical system, in which the consumers switch between providers based on online reviews, as the time-varying online review system is continuously updated by the consumer fluxes. We apply the model to various network structures, capturing providers' accessibility: i) random, canonical networks and ii) a real-world network of medical doctors in New York City. We examine the emerging correlations and causal relationships between the success of providers and the topological properties of the networks. Across a wide range of networks of varying size, we consistently find that online reviews have an important role in providers' success. The satisfaction of the consumers in the online review systems, together with the market share, influences consumer fluxes between providers and the overall quality of service experienced by consumers. The study of the network of doctors reveals some causal mechanisms in the decision-making processes, with the doctor's success impacting on the providers' quality of service and the consumer fluxes.**

***Index Terms*—Complex systems, decision-making, Markov chain, online reviews, urban data.**

I. Introduction

ONLINE review systems (ORSs) are crucial for active consumers who search for evaluations and opinions on products and services [1], [2]. These platforms provide an important resource for consumers seeking to make informed choices that align with their individual preferences and needs. Cheung et al. [2] and Alkalbani et al. [3] found that the motivations, attitudes, and behaviors of consumers to provide reviews on an ORS are also important to investigate. People write reviews to express their opinions and help others make informed decisions [4], [5], [6]. ORS platforms provide valuable insights that enable shoppers to make informed decisions regarding their purchases or subscriptions, based on the perceived quality of products or services provided in the market [7]. They allow consumers to access a wealth of information about various products and services. The performance of services or providers is also a determinant of consumer satisfaction [8], [9]. To satisfy the needs of consumers, providers have to maintain a high quality of service in a competitive environment, since unsatisfied consumers can easily switch to a different service provider.

While ORSs offer information about any provider, even those remotely located from the consumers, location and proximity remain important factors that influence consumer choices [10]. For example, in healthcare, patients may prefer to choose family doctors who are located near their homes or workplaces, as this reduces transportation costs and provides easier/quicker access in case of medical issues or concerns. Therefore, ORSs that allow users to search for products or services based on their geographical location can be particularly helpful where the options are plentiful [11], [12]. Agencies or industries are often geographically clustered with respect to socioeconomic and demographic factors [13], so that many businesses are at a high risk of losing market share due to the intense competition [14], [15].

Understanding the importance of ORSs requires studying a complex system of intricate social and economic interactions describing how humans make decisions with regard to geographical and socioeconomic factors over online social networks. Computational models are widely used for explaining human choices. For example, models have been developed to examine opinion dynamics and the coexistence and competition of opposing options. Quattrociocchi et al. [16] put forward a model to study opinion dynamics accounting for the coexistence of media communication and social influence. Antonopoulos and Shang [17] studied the information diffusion and opinion dynamics over multiplex networks. Hudson and Khamfroush [18] proposed an optimization algorithm to maximize information diffusion in community-based online social networks. Friedkin et al. [19] modeled a network-based belief system with heterogeneous nodes (beliefs with different certainties), where multiple statements about a topic exist. These models only consider agents exchanging opinions over social networks about a single topic that can prompt consensus or polarity.

In the real world, a consumer has to form an opinion and make a selection from multiple services or providers.

Manuscript received 14 August 2023; revised 19 November 2023; accepted 27 December 2023. Date of publication 4 January 2024; date of current version 30 April 2024. This work was supported by National Science Foundation under Grants ECCS-1928614 and DUE-2129076. Recommended for acceptance by Dr. Fan Wu. *(Corresponding author: Maurizio Porfiri.)*

Tian Gan is with the Center for Urban Science and Progress and Department of Mechanical and Aerospace Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA (e-mail: tg2426@nyu.edu).

Rishita Das is with the Center for Urban Science and Progress and Department of Mechanical and Aerospace Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA, and also with the Department of Aerospace Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: rd3154@nyu.edu).

Maurizio Porfiri is with the Center for Urban Science and Progress, Department of Mechanical and Aerospace Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA, and also with the Department of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA (e-mail: mporfiri@nyu.edu).

Digital Object Identifier 10.1109/TNSE.2023.3348654

Therefore, it is important to model the multiple-choice behavior of humans to understand consumer-provider systems fully. In this regard, Hausman et al. [20] proposed a logit-based model for consumers that considers multiple services and the utilities of all these services. Utility is a measure of an individual's pleasure or happiness as a consumer of a service. In economics, an individual's preference between alternatives is determined by the expected value of the service provided by each alternative for the individual. For example, in the decision-making process of selecting from multiple residential locations, a multinomial choices model was proposed by Kim et al. [21]. While existing choice models focus on the utility function of individual decisions, the accessibility of a product or service is a key aspect of decision-making.

In this work, we address the question of how consumers' reviews and providers' accessibility affect consumers' decision-making and the overall consumer-provider dynamics. We posit a network dynamical system to gain an improved understanding of consumer-provider relationships and consumers' decision-making. Within our modeling framework, we construct a state-dependent flow network of providers in which consumers flow from one provider to another. The network topology reflects the accessibility of the providers to the consumers, whereby, in real-world scenarios, accessibility depends on geographic [22], [23] and socioeconomic [24], [25] factors. Critical to the proposed model is the presence of an ORS, which influences the selection process that underpins the consumer fluxes between providers. Before availing a service, consumers select a preferred provider from the pool of accessible providers based on their online ratings/reviews. After the service, they express their opinion about their providers on the ORS, thereby adding to the existing ratings/reviews. The performance of the provider and their customer load may have an impact on the wait times and the quality of their service, thus affecting consumers' satisfaction [26].

Our network model can be adapted to a number of real-world scenarios involving people's selection of providers, such as gym membership, car-sharing and bike-sharing subscriptions, academic publishing, and primary care physicians. To detail the emerging selection dynamics, we consider two canonical network structures for providers' accessibility (Barabási-Albert and Erdös-Rényi networks). As a demonstration of the model to a real-world problem, we investigate a case study in healthcare, which is presently facing dire challenges due to an overload of existing providers [27]. The consequences of this overload include compromised quality of care, extended wait times for appointments with doctors, and even delayed diagnoses for patients, along with extreme competition between providers [28]. We examine medical care in New York City (NYC), taking into account accessibility and online reviews of doctors. We establish causal relationships among the system variables, including the competition faced by the doctors, the quality of their service, their ability to gain patients, and their patients' satisfaction.

The main contribution of our work is a generalized network-based model for a consumer-provider system where the consumers make decisions on selecting a provider based on the providers' online reviews and accessibility. The model captures the dynamics of consumers' selection and satisfaction among a network of providers. Our findings significantly enhance the understanding of consumers' decision-making process regarding the selection of providers, as a valuable basis for informing policy development in a variety of fields, such as business, healthcare, and publisher aspects.

The remaining paper is organized as follows. In Section II, we present our modeling framework. In Section III, we apply our model to canonical networks. Section IV is devoted to a case study of our model in the healthcare sector and to detail analysis of the inner workings of system dynamics through causal inference. Section V summarizes our main conclusions and outlines avenues for future research.

II. Model Formulation

A. Problem Setting

The selection of a provider is influenced by various factors, including distance from one's residence, wait times for appointments, satisfaction levels with the provider, transportation, and access to the provider's practice, as well as socioeconomic and demographic characteristics [29], [30].

In Fig. 1(a), we illustrate the process of provider selection performed by each consumer. When a consumer seeks a service or a provider, they consult an ORS platform to learn about other consumers' evaluations of the providers' performance and make a selection accordingly [31]. After meeting with the chosen provider, the consumer may provide an online rating and a review about their own experience on the ORS [1]. Dissatisfied consumers are likely to leave a negative review and switch to another provider, selecting one from the remaining pool of providers based on the ORS, and they will continue doing so until they find a satisfactory provider [32]. Conversely, satisfied consumers are likely to leave positive reviews and continue with their selected providers [32].

With respect to online reviews, previous research [33], [34] suggests that a relatively small proportion of consumers write reviews and that these reviews are typically biased. Thus, we consider that only a proportion of satisfied or dissatisfied consumers give an online rating and write down a textual review after changing their provider. In fact, an even smaller percentage of dissatisfied consumers is assumed to provide a negative online review [4]. Consumers changing to a new provider may also give their feedback if they decide to continue with this provider.

B. Capacity and Flow Networks

In our consumer-provider system, we have a set of N providers $\mathcal{P} = \{P_1, \ldots, P_i, \ldots, P_N\}$ and a set of M consumers $\mathcal{C} = \{C_1, \ldots, C_l, \ldots, C_M\}$. To analyze the impact of online reviews and accessibility factors on the selection process, we posit a network-based dynamical system. Provider P_i is accessible to a fraction of the consumers, some of whom could access other providers in the network. The undirected, weighted capacity network $\mathcal{G}_\mathrm{c}$ over the vertex set $\mathcal{P}$ models the accessibility of providers. The edge set of such a network, $\mathcal{E}_\mathrm{c}$, identifies pairs of providers who have mutually accessible consumers. The weight F_{ij} of the edge between providers P_i and P_j is given by the

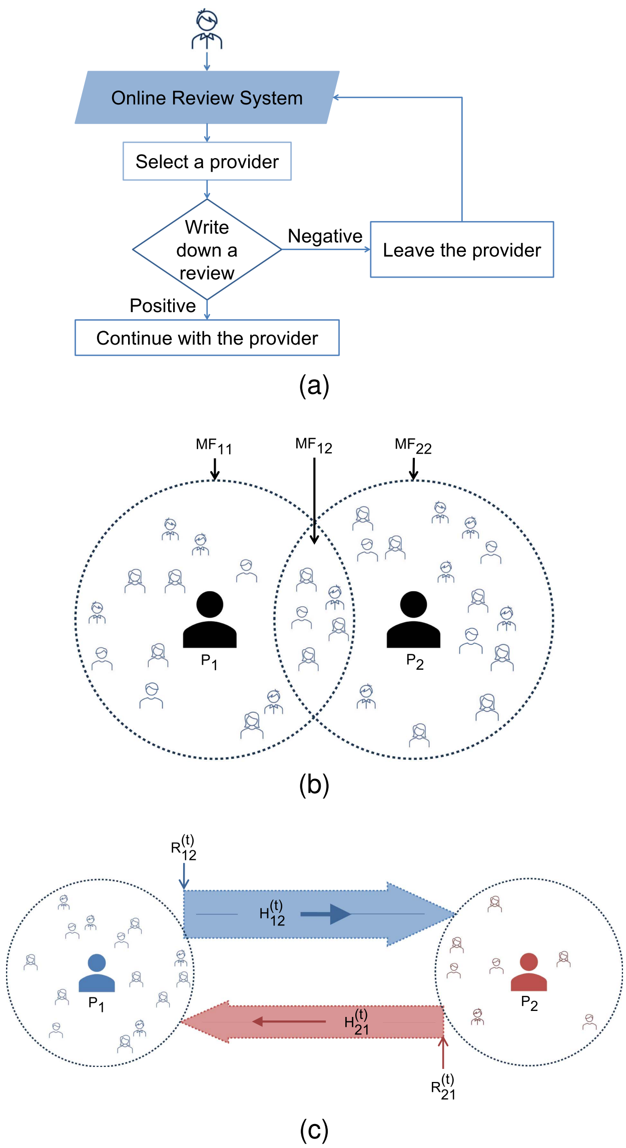

Fig. 1. Illustration of our modeling framework. (a) A schematic representation of the selection process exercised by a consumer. (b) An example of a consumer-provider system with two providers $\mathcal{P} = \{P_1, P_2\}$ and M consumers. Consumers who have access to a provider are marked within the corresponding circle surrounding the provider. The consumers in the intersection of the two circles have access to both providers. F_{12} is the fraction of consumers who have access to both providers. F_{11} and F_{22} represent the fractions of consumers who have access to providers P_1 and P_2, respectively. This system represents a capacity network $\mathcal{G}_c$ of two nodes, P_1 and P_2, connected by an edge of weight F_{12} and consisting of self-loops F_{11} and F_{22}, respectively. (c) An example of a consumer-provider system with two providers are accessible to $M(R_{12}^{(t)} + R_{21}^{(t)})$ consumers. Consumers who are subscribed to a provider are marked within the corresponding circle surrounding the provider (for example, MR_{12} is the number of consumers who can access both providers and are currently subscribed to P_1). $H_{12}^{(t)}$ is the consumers flux from P_1 to P_2 at t.

fraction of consumers to whom both P_i and P_j are accessible. For simplicity, we assume that a consumer has at most access to two providers. We consider the presence of self-loops, whereby F_{ii} represents the fraction of consumers with access to P_i; by definition $F_{ij} \leq F_{ii}$ for any $(i,j) \in \mathcal{E}_c$ and $F_{ii} = \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} F_{ij} + \hat{F}_{ii}$ where $\mathcal{N}_c(i)$ is the set of neighbors of provider i, excluding themselves ($\mathcal{N}_c(i) = j \in 1,\ldots,N$ such that $(i,j) \in \mathcal{E}_c$), and $\hat{F}_{ii}$ is the fraction of consumers who can only access provider P_i. In Fig. 1(b), we show an example for two providers.

To track the flow of consumers between providers, we also construct a directed, time-varying, weighted flow network $\mathcal{G}^{(t)}$ of providers with the same vertex set, $\mathcal{P}$. The edge set is also the same as the capacity network, $\mathcal{E}_c$, but, this time, we use directed edges in the network to differentiate the two fluxes between providers who share some consumers. We define a weight (potentially zero) for any of the directed edges in the capacity network to describe the dynamics of provider selection over time ($t \in \mathbb{Z}_+ \cup \{0\}$). The extent of the consumer flux is captured by the weight $H_{ij}^{(t)}$, which is the fraction of consumers switching from P_i to P_j at t. Self-loops of the type $H_{ii}^{(t)}$ identify consumers who are dissatisfied but still remain with provider i at t. The quantity $\sum_{j\in\mathcal{N}_c(i)} H_{ij}^{(t)}$ represents the fraction of consumers who are dissatisfied with provider P_i, but may still remain with them. An example of this flow network with two providers is shown in Fig. 1(c).

C. System Dynamics

We model the selection process as a discrete-time Markov process within the flow network $\mathcal{G}^{(t)}$,

$$X_i^{(t+1)} = X_i^{(t)} + \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} (H_{ji}^{(t)} - H_{ij}^{(t)}), \tag{1}$$

for $i = 1\ldots,N$. Here, $X_i^{(t)}$ is the fraction of consumers who are subscribed to provider P_i. On the right-hand side of (1), $\sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ij}^{(t)}$ is the loss of consumers experienced by provider P_i (which must be less than $X_i^{(t)}$) and $\sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ji}^{(t)}$ is the gain in consumers of P_i. The model is initialized at $t = 0$ with initial conditions that satisfy $\sum_{i=1}^N X_i^{(0)} = 1$.

Property 1: For system (1), one can verify that $\sum_{i=1}^N X_i^{(t)} = 1$ for any t, by summing both sides of (1) over i.

Lemma 1: System (1) begets a well-defined time-evolution for the fraction of consumers, so that model dynamics will always lead to solutions in [0, 1] for initial conditions in [0, 1].

Proof: The claim can be verified by bounding from below the right-hand side of (1) by $X_i^{(t)} - \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ij}^{(t)}$. By noting that $\sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ij}^{(t)}$ is the outward flux from provider P_i, which must be less than their fraction of consumers, $X_i^{(t+1)} \geq 0$ for any t. Given that $\sum_{i=1}^N X_i^{(t)} = 1$ from Property 1, all of the $X_i^{(t)}$s' are fractions. ■

The consumer fluxes vary stochastically in time according to two dependent events: being dissatisfied with the current provider and selecting a provider in $\mathcal{P}$ upon being dissatisfied with the current provider. The probability of a generic consumer to switch from provider P_i to P_j at t is the product of the probability $f(X_i^{(t)})$ that a consumer is dissatisfied with their provider P_i and the conditional probability $\sigma_{j|i}^{(t)}$ that the consumer selects P_j given that they are dissatisfied with P_i. We consider the case in which $i = j$ as well, so that a consumer is dissatisfied with their provider and yet they can remain with them. By construction, we must satisfy $\sum_{n\in\{i,j\}} \sigma_{n|i}^{(t)} = 1$ so that a consumer who is dissatisfied with P_i either chooses a new provider P_j or stays with P_i.

Accounting for the fact that only a fraction of consumers who have access to both P_i and P_j are currently with P_i, the consumer flux from P_i to P_j at t is given by

$$H_{ij}^{(t)} = R_{ij}^{(t)} f(X_i^{(t)}) \sigma_{j|i}^{(t)}, \tag{2}$$

where $R_{ij}^{(t)}$ is the fraction of consumers who have access to P_i and P_j and are subscribed to provider P_i at t, see Fig. 1(c). By definition,

$$X_i^{(t)} - \hat{F}_{ii} = \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} R_{ij}^{(t)}, \tag{3}$$

where $X_i^{(t)} - \hat{F}_{ii}$ is the fraction of consumers who are currently subscribed to P_i and can move from P_i. In the case $i = j$, $R_{ii}^{(t)}$ is simply $X_i^{(t)}$. For $i \neq j$, this quantity varies in time according to the consumer flux between P_i and P_j,

$$R_{ij}^{(t+1)} = R_{ij}^{(t)} - H_{ij}^{(t)} + H_{ji}^{(t)}, \tag{4}$$

where the second term on the right-hand-side captures reduced capacity due to flux from P_i to P_j and the third increased capacity due to flux from P_j to P_i. Initial conditions for the residuals are such that $R_{ij}^{(0)} \geq 0$ and $R_{ij}^{(0)} + R_{ji}^{(0)} = F_{ij}$ for $(i, j) \in \mathcal{E}_c$ and $i \neq j$.

Remark 1: Equation (4) is based on the assumption that a consumer has at most access to two providers; should one contemplate releasing this assumption, this equation would not be valid as there could be consumers who not only have access to providers P_i and P_j, but also to P_k ($k \neq i, j$). A remedial strategy is to use tensorial quantities, defining higher-order variables like $R_{ijk}^{(t)}$ in the vein of hypergraphs [35].

Property 2: By swapping the indices in (4), we have that $R_{ij}^{(t)} + R_{ji}^{(t)}$ remains constant in time for for $(i, j) \in \mathcal{E}_c$ and $i \neq j$. Therefore, $R_{ij}^{(0)} + R_{ji}^{(0)} = F_{ij}$ implies that $R_{ij}^{(t)} + R_{ji}^{(t)} = F_{ij}$ at all times.

Based on this property, we can demonstrate intuitive, individual bounds for $R_{ij}^{(t)}$ and $R_{ji}^{(t)}$ as follows.

Lemma 2: The dynamics of the residuals according to (4) is such that $0 \leq R_{ij}^{(t)} \leq F_{ij}$ for $(i, j) \in \mathcal{E}_c$ and $i \neq j$.

Proof: We first proceed by induction to show that the residuals are non-negative. By bounding from below the right-hand side of (4) by $R_{ij}^{(t)} - H_{ij}^{(t)}$, one can prove that $R_{ij}^{(t+1)} \geq 0$ provided that $R_{ij}^{(t)} \geq 0$ (since $R_{ij}^{(t)} - H_{ij}^{(t)} \geq 0$ according to (2)). Given that $R_{ij}^{(t)}$ and $R_{ji}^{(t)}$ are non-negative and $R_{ij}^{(t)} + R_{ji}^{(t)} = F_{ij}$ based on Property 2, we conclude that $R_{ij}^{(t)} \leq F_{ij}$. ∎

Using Lemma 2, we can prove an important claim on the evolution of the fractions of consumers.

Lemma 3: System (1) and residuals in (4) evolve in time so that $X_i^{(t)} \in [\hat{F}_{ii}, F_{ii}]$.

Proof: We recall that $\hat{F}_{ii} = F_{ii} - \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} F_{ij}$. Using (1), (3), and (4), one can show that $X_i^{(t+1)} = X_i^{(t)} - \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} R_{ij}^{(t)} + \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} R_{ij}^{(t+1)} = F_{ii} - \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} F_{ij} + \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} R_{ij}^{(t+1)}$. Given that $R_{ij}^{(t+1)} \in [0, F_{ij}]$ based on Property 2, it follows that $X_i^{(t+1)} \in [\hat{F}_{ii}, F_{ii}]$. ∎

Variables $H_{ij}^{(t)}$ capture all consumer fluxes between providers, ultimately describing the competition for consumers experienced by the providers. As time progresses, provider P_i will compete for consumers with a varying set of providers. The competition experienced by a provider is measured by the influx $\sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ji}^{(t)}$ and the outflux $\sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ij}^{(t)}$. At the steady state, when the fraction of consumers in (1) plateaus, the influx balances the outflux for each provider. We collate these quantities in the vector $\bar{k}$, whose generic entry is the influx (outflux) of provider P_i at the steady state. Such a vector can be viewed as a weighted degree of the flow network at the steady state, upon discounting self-loops.

To model the probability that a consumer is dissatisfied with their current provider P_i, we consider the following factors: the decline in the provider's efficiency due to increasing consumers to cater to [36] and the provider's performance towards each consumer [37]. We formulate this as follows:

$$f(X_i^{(t)}) = \begin{cases} \frac{F_{ii}}{1+e^{-W(X_i^{(t)}-E)}} + \frac{1}{L_i} & \text{if } \frac{F_{ii}}{1+e^{-W(X_i^{(t)}-E)}} + \frac{1}{L_i} \leq 1; \\ 1 & \text{otherwise,} \end{cases} \tag{5}$$

where the logistic function $\frac{F_{ii}}{1+e^{-W(X_i^{(t)}-E)}}$ captures the inefficiency of provider P_i due to the increase in demand and $\frac{1}{L_i}$ their inadequacy as experienced by the consumers. Parameters $W > 0$, $E > 0$, and F_{ii} define the shape of the logistic function, and $L_i > 0$ the provider's performance.

We rely on the online reviews of the providers to model the probability of selecting a provider over other providers ($\sigma_{i|j}^{(t)}$). To simplify the model, we group online ratings in the ORS into a set of positive and a set of negative ratings. The overall online satisfaction rate $S_i^{(t)}$ for each provider P_i is

$$S_i^{(t)} = \frac{S_{i,+}^{(t)}}{S_{i,+}^{(t)} + S_{i,-}^{(t)}} \in [0, 1] \tag{6}$$

where $S_{i,+}^{(t)}$ and $S_{i,-}^{(t)}$ are the number of positive ratings and that of negative ratings for the provider P_i, respectively. Recent work [38] has shown that trustworthiness of ratings is important for the consumers' selection process. When the online satisfaction rate $S_i^{(t)}$ is high, the trust in the rating depends on the number of textual reviews $T_i^{(t)}$ about the provider P_i. Otherwise,

$T_i^{(t)}$ has no effect on how trustworthy the rating is. We define trustworthiness $\tau_i^{(t)}$ as

$$\tau_i^{(t)} = \begin{cases} \mathrm{sigmoid}(T_i^{(t)}), & \text{if } S_i^{(t)} > \frac{1}{2}; \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

To compare preference for the two accessible providers, we assume the decision of choosing P_i is based on the expected utility function $V_{i|j}$ of moving from P_j to P_i at t,

$$V_{i|j}^{(t)} = V_i - V_j = \tau_i^{(t)} S_i^{(t)} - \tau_j^{(t)} S_j^{(t)} \tag{8}$$

where $V_i^{(t)}$, the utility function of provider P_i, is given by the overall online satisfaction rate $S_i^{(t)}$ in (6) and trustworthiness $\tau_i^{(t)}$ in (7), for P_i. Thus, the probability that a consumer selects provider P_i is in the form of a multi-nomial choice model of P_i's expected utility at t, given by a softmax function,

$$\sigma_{i|j}^{(t)} = \frac{e^{V_{i|j}^{(t)}}}{\sum_{n\in\{i,j\}} e^{V_{n|j}^{(t)}}} = \frac{e^{\tau_i^{(t)} S_i^{(t)} - \tau_j^{(t)} S_j^{(t)}}}{\sum_{n\in\{i,j\}} e^{\tau_n^{(t)} S_n^{(t)} - \tau_j^{(t)} S_j^{(t)}}}. \tag{9}$$

D. Time Evolution of the ORS

The time-varying ORS is updated by the consumers who switch their providers. The consumer who switches to a new provider and is satisfied with the experience is likely to leave a positive rating and review. Meanwhile, the consumer who switches from a provider (to another provider) is likely to provide a negative rating and review about the provider they left. In reality, not all consumers are willing to express their level of satisfaction [39]. The number of positive and negative ratings of a provider P_i at t is modeled by

$$S_{i,+}^{(t+1)} = S_{i,+}^{(t)} + \alpha \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ji}^{(t)} Q_i^{(t)}, \tag{10a}$$

$$S_{i,-}^{(t+1)} = S_{i,-}^{(t)} + \beta \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ij}^{(t)}. \tag{10b}$$

where parameters $\alpha \in [0,1]$ and $\beta \in [0,1]$ represent the proportion of consumers who are willing to give a positive and a negative rating, respectively. Here, $Q_i^{(t)}$ is the service quality of provider P_i, given by

$$Q_i^{(t)} = 1 - f(X_i^{(t)}), \tag{11}$$

which also represents the probability that a consumer is satisfied with the service of provider P_i at t. Equation (10a) captures the fraction of consumers who switch to P_i and are satisfied with this new provider P_i (giving a positive rating). Equation (10b) indicates that P_i's negative ratings are updated by the proportion of consumers who switch from provider P_i to any other provider.

Besides, we consider that some of the reviewers would like to write down a textual review. The number of textual reviews

Algorithm 1: System Dynamics on an N-Node Network.

Initialization: Set parameters W, E, α, β, and γ for (5), (10a), (10b), and (12). Begin at $t = 0$ with $\mathcal{G}^{(0)}$, $X_i^{(0)}, S_{i,+}^{(0)}, S_{i,-}^{(0)}, T_i^{(0)}$, and $R_{ij}^{(0)}$, $\forall i, j \in \mathcal{N}$,

while ($t <$ max simulation time) **do**
  for $i = 1$ to N **do**
    $S_i^{(t)} \leftarrow S_{i,+}^{(t)}, S_{i,-}^{(t)}$ via (6)
    for $j = 1$ to N **do**
      if $j \neq i$ **then**
        $H_{ij}^{(t)} \leftarrow X_i^{(t)}, R_{ij}^{(t)}, S_i^{(t)}, T_i^{(t)}$ based on (2), (5), (7), and (9)
      end if
    end for
    $X_i^{(t+1)} \leftarrow X_i^{(t)}$ via (1)
    $R_{ij}^{(t+1)} \leftarrow R_{ij}^{(t)}, H_{ij}^{(t)}$, and $H_{ji}^{(t)}$ via (4)
    $S_{i,+}^{(t+1)} \leftarrow S_{i,+}^{(t+1)}$, and $H_{ji}^{(t)}$ via (10a), and (11)
    $S_{i,-}^{(t+1)} \leftarrow S_{i,-}^{(t+1)}$, and $H_{ij}^{(t)}$ via (10b)
    $T_i^{(t+1)} \leftarrow T_i^{(t)}$ via (12)
  end for
  $\mathcal{G}^{(t+1)} \leftarrow \mathcal{G}^{(t)}, H_{ij}^{(t)}$, and $H_{ji}^{(t)}$
  $t \leftarrow t + 1$
end while

is updated as follows:

$$T_i^{(t+1)} = T_i^{(t)} + \gamma \left[\alpha \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ji}^{(t)} Q_i^{(t)} + \beta \sum_{\substack{j\in\mathcal{N}_c(i)\\ j\neq i}} H_{ij}^{(t)} \right] \tag{12}$$

where $T_i^{(t)}$ is increased by the consumers who not only give their ratings but also textual reviews to provider P_i, and $\gamma \in [0,1]$ is the proportion of the reviewers providing a rating who are willing to write down a textual review as well.

In summary, providers $\mathcal{P}$ at t have fractions of currently subscribed consumers $X^{(t)} = (X_1^{(t)}, \ldots, X_N^{(t)})^\top$, online satisfaction rates $S^{(t)} = (S_1^{(t)}, \ldots, S_N^{(t)})^\top$, and numbers of textual reviews $T^{(t)} = (T_1^{(t)}, \ldots, T_N^{(t)})^\top$. Thus, each node P_i has a vector of current-state $[X_i^{(t)}, S_i^{(t)}, T_i^{(t)}]$, making the network dynamics vectorial. An accompanying pseudo-code is presented in Algorithm 1 to outline the main steps for the implementation of the model.

III. Results on Canonical Networks

Here, we apply our model to different instances of flow networks. We generate several random graphs for the capacity network of providers $\mathcal{G}_c$ and study variations of the consumer fluxes in the provider networks of providers $\mathcal{G}^{(t)}$.

We study a series of systems with $N \in [100, 1,000]$ providers and $M = 5,000$ consumers. First, we initialize the capacity network $\mathcal{G}_c$ as an unweighted random graph drawn from a given canonical model; here and in what follows, we use superscript "unw" to refer to this network, $\mathcal{G}_{\mathrm{c}}^{(\mathrm{unw})}$. Then, we randomly allocate the consumers $\mathcal{C} = \{C_1, \ldots, C_M\}$ to the providers in the network so that each provider ($P_i, i = 1, \ldots, N$) has a set of accessible consumers (F_{ii}) and connected providers have a set of mutually accessible consumers (F_{ij}). This process produces the weights of the capacity network, thereby transforming $\mathcal{G}_{\mathrm{c}}^{(\mathrm{unw})}$ into $\mathcal{G}_{\mathrm{c}}$. By definition, $\hat{F}_{ii} = F_{ii} - \sum_{\substack{j\in\mathcal{N}_{\mathrm{c}}(i)\\ j\neq i}} F_{ij}$ is the fraction of consumers who can only access the single provider P_i.

The initial number of positive and negative ratings ($S_{i,+}^{(0)}$ and $S_{i,-}^{(0)}$) and textual reviews ($T_i^{(0)}$) of provider P_i are assigned based on a uniform random distribution in [1, 100], for $i = 1, \ldots, N$. From (6), the initial online satisfaction rate $S_i^{(0)}$ of P_i is computed. Using the initial conditions of the ORS ($S^{(0)} = (S_1^{(0)}, \ldots, S_N^{(0)})^\top$ and $T^{(0)} = (T_1^{(0)}, \ldots, T_N^{(0)})^\top$) and the set of accessible consumers of the providers, we assign the consumers to the providers such that providers with higher initial satisfaction rate and more number of textual reviews are likely to have higher initial number of consumers. Specifically, we compute the probability that a consumer selects provider P_i using (9), and we create a related cumulative distribution function (CDF). We use the CDF to assign consumers to providers via the inverse transform sampling method [40]. In this manner, we obtain the initial fractions of subscribed consumers of the providers $X^{(0)} = (X_1^{(0)}, \ldots, X_N^{(0)})^\top$ and the residuals $R_{ij}^{(0)}, \forall i, j = 1, \ldots, N$, thereby forming the initial flow network $\mathcal{G}^{(0)}$.

To simulate the logistic function in (5), we set $W = 0.08$ and $E = 50$. Parameters α, β, and γ are set as 0.2, 0.05, and 0.5, respectively. These settings approximate a real-world scenario where a smaller percentage of dissatisfied consumers is assumed to provide a negative review than the percentage of satisfied consumers who provide positive reviews [4].

The simulations consist of numerically propagating (1), (10), and (12), for 3,000 time steps to reach an equilibrium. Throughout our numerical study, we always register that solutions converge to a steady state so that all providers reach a steady number of consumers by the end of the simulation. For some providers, the fraction of consumers gradually decreases or increases and then converges to a steady state. In contrast, some providers find it difficult to attract consumers and remain at a nearly unchanged fraction of consumers over the entire duration. The steady-state values that $X^{(t)}$, $S^{(t)}$, and $Q^{(t)}$ converge to, are denoted by $\bar{X}$, $\bar{S}$, and $\bar{Q}$, respectively. These quantities are examined alongside $\bar{k}$, capturing the steady-state consumer fluxes.

The fraction of consumers currently subscribed to a provider P_i is also called the market share of the provider [41]. The Herfindahl-Hirschman index (HHI index) gives the competition for the provider networks,

$$\mathrm{HHI} = \|\bar{X}\|^2, \tag{13}$$

where $\|\bar{X}\|$ is the Euclidean norm of vector $\bar{X}$. An increase in the HHI index represents a decrease in the market competition [42]. The HHI index is bounded between $\frac{1}{N}$ and 1.

A. Barabási-Albert Model

First, we examine networks generated by the Barabási-Albert (BA) model [43] that creates random scale-free networks with a power-law degree distribution. To form a BA network of N nodes, we start with a small star network of $m \in \{1, 2, \ldots, 9, 10\}$ nodes and then we introduce $N - m$ new nodes based on the preferential attachment rule [44]. For N sufficiently large, the mean degree of $\mathcal{G}_{\mathrm{c}}^{(\mathrm{unw})}$ is approximately $2m$ [43].

We begin the analysis of the BA model by studying the distribution of $\bar{X}$ for $N \in [100, 1,000]$, see Fig. 2(a). Our results suggest that the steady-state market share of the providers is more concentrated around the median for larger markets, see Fig. 2(a). In other words, there is a lower variance in the success of the providers to attract and retain consumers if the system has a larger number of providers. To delve into the competition among providers, we study the steady-state consumer fluxes, $\bar{k}$, see Fig. 2(b). The median value of $\bar{k}$ gradually decreases with increasing network sizes, until reaching a plateau slightly above zero. For BA networks of different sizes N, we plot the corresponding HHI indices in Fig. 2(c). As the size of the market increases, HHI decreases, indicating higher competition among providers. A lower HHI could be explained by the fact that providers who are clustering together might have a similar market share, as the same set of consumers would be able to access them.

Next, we detail the relationship between the performance L of providers, consumers' satisfaction $\bar{S}$, market share $\bar{X}$, and steady-state consumer fluxes $\bar{k}$ of the providers at the steady state, see Fig. 2(d) and (e). Results show that the providers with higher L_i are more likely to attract consumers and obtain higher ratings. In these figures, we divide the providers into low- ($\bar{k}_i \leq \langle\bar{k}\rangle$) and high-degree ($\bar{k}_i > \langle\bar{k}\rangle$) providers, based on the average value $\langle\bar{k}\rangle$. The classification reveals that high-degree providers have more consumers than low-degree providers. This is likely because having more competitors amplifies the possibility of competing for consumers with other providers so that providers can attract and retain more consumers, irrespective of their performance. Providers' performance, however, is positively related to the consumers' satisfaction. The relationship between the market share $\bar{X}$ and the steady-state consumer fluxes $\bar{k}$ of the flow network for $N \in [100, 1,000]$ and $m \in [1, 10]$ is illustrated in Fig. 2(f). Aggregated results show that providers with high steady-state consumer fluxes could attain more consumers on average. Meanwhile, for providers with lower weighted degrees, there exists more variability in their steady-state market share.

B. Erdős–Rényi Model

Similar to the previous section, the Erdős-Rényi (ER) model is used to generate initial topology structures of the capacity network $\mathcal{G}_{\mathrm{c}}^{(\mathrm{unw})}$ of providers and to investigate the consumer fluxes in such a network. To construct an ER network with N nodes, all edges are independently connected to each of two nodes with a fixed edge probability p. The mean degree of the

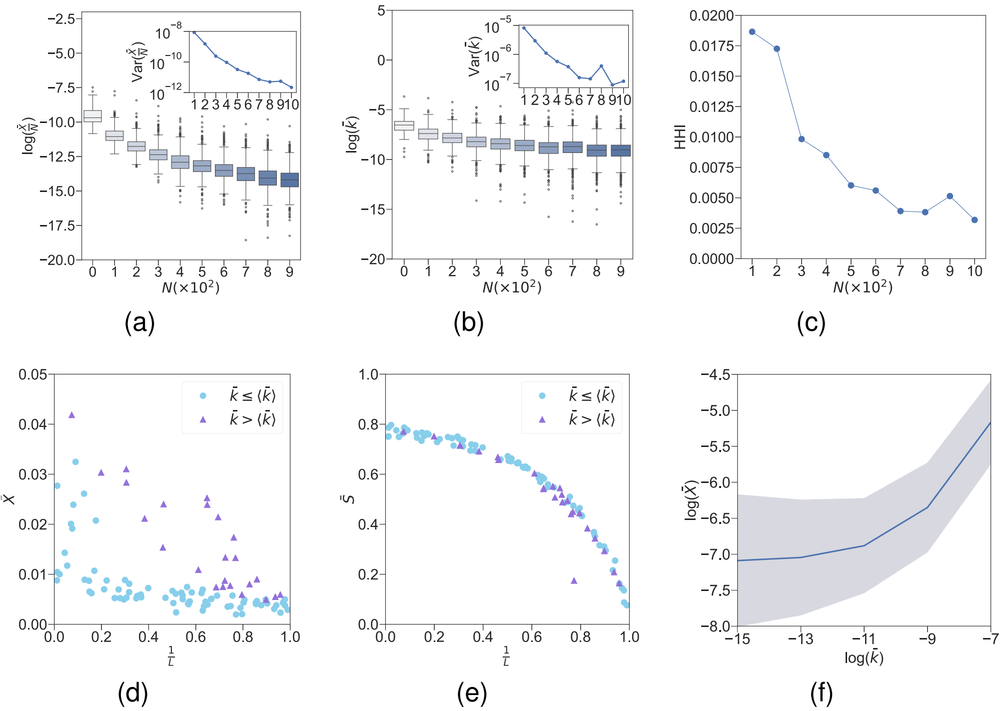

Fig. 2. Model prediction for BA networks. (a) Box plots for the distribution of steady-state market share ($\bar{X}$) for $N \in [100, 1,000]$ nodes and $m = 7$ edges for the preferential attachment, with variances in the inset. (b) Box plots for the distribution of weighted degrees of providers ($\bar{k}$) for $N \in [100, 1,000]$ and $m = 7$, with variances in the inset. On each number, the central mark is the median, and the bottom and top edges are the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data that are not outliers, and outliers are plotted individually using dots. (c) Scatter plot of the HHI for BA networks generated by $N \in [100, 1,000]$ and $m = 7$. (d)-(e) Scatter plots of (d) the inverse of performance $\left(\frac{1}{L} = [\frac{1}{L_1}, \ldots, \frac{1}{L_N}]^\top\right)$ and steady-state market share $\bar{X}$; and (e) the inverse of performance and steady-state consumers' satisfaction $\bar{S}$ for all providers in the system. The providers are classified by the average consumer fluxes $\langle \bar{k} \rangle$ of the steady-state provider network. In this case, $\mathcal{G}_c^{(\text{unw})}$ is a BA network generated with $N = 100, m = 7$. (f) Log-scale error plot of the steady-state market share $\bar{X}$ for different steady-state consumer fluxes $\bar{k}$ for BA networks with $N \in [100, 1,000]$ and $m \in [1, 10]$. The solid line and the shaded area represent the mean and the standard deviation, respectively, of the distributions of $\bar{X}$ conditioned on $\bar{k}$, organized into five equally spaced intervals.

generated ER network is approximately Np, which is a function of the number of nodes N and the rewiring probability p [43].

The capacity network $\mathcal{G}_c^{(\text{unw})}$ generated by ER networks with $N \in [100, 1,000]$ have the same mean degree as the BA networks generated with $m \in [1, 10]$ edges based on preferential attachment. Thus, we compare the simulation results on BA and ER models where their capacity network $\mathcal{G}_c^{(\text{unw})}$ has the same mean degree but different network structure.

Overall, the analysis of ER networks is in strong qualitative agreement with the one of BA networks. Differences between the two canonical networks are mostly in the range of the observed trends and variance of the statistical observation. Specifically, for the same mean degree as the capacity network $\mathcal{G}_c^{(\text{unw})}$ generated by the BA model, the market share $\bar{X}$ and the steady-state consumer fluxes $\bar{k}$ display equivalent variation with N, (shown in Fig. 3(a) and (b), respectively,) to the results of BA networks in Fig. 2(a) and (b). The main difference between ER and BA networks is the statistical distributions, which, as one should expect, have stronger outliers in the case of BA networks. Similar to Fig. 2(c), the HHI in Fig. 3(c) declines as the market size N increases, indicating increasing competition among providers. The variation of market share $\bar{X}$ and steady-state consumer fluxes $\bar{k}$ with performance L are plotted in Fig. 3(d), (e), and (f). Providers with high performance could obtain more consumers or higher ratings. The assessment of the competition experienced by providers, as determined by their steady-state consumer fluxes with respect to the average value $\langle \bar{k} \rangle$, reveals that high-degree providers are likely to compete with more neighboring providers to draw consumers irrespective of their performance. The relationship between $\bar{k}$ and $\bar{X}$ of the ER network with $N \in [100, 1,000]$ and $m \in [1, 10]$ in Fig. 3(f) suggests that providers experiencing large consumer fluxes are likely to attain more consumers, similar to the BA networks in Fig. 2(f).

IV. Case Study: Doctor Selection in NYC

A. *Database of Doctors*

Here, we implement our model on a healthcare system to detail patients' decision-making process in the selection of doctors and study the link between the inefficiency of the service and limited access to healthcare of patients. We collect data on 487 family doctors, or primary care physicians (PCPs), in NYC, including their names, geographical coordinates of their practice, online ratings, online textual reviews, and types of health insurance accepted. We use a web scraper called BeautifulSoup toolbox [45] to collect the data on online ratings and textual reviews about the doctors from the website www.healthgrades.com.

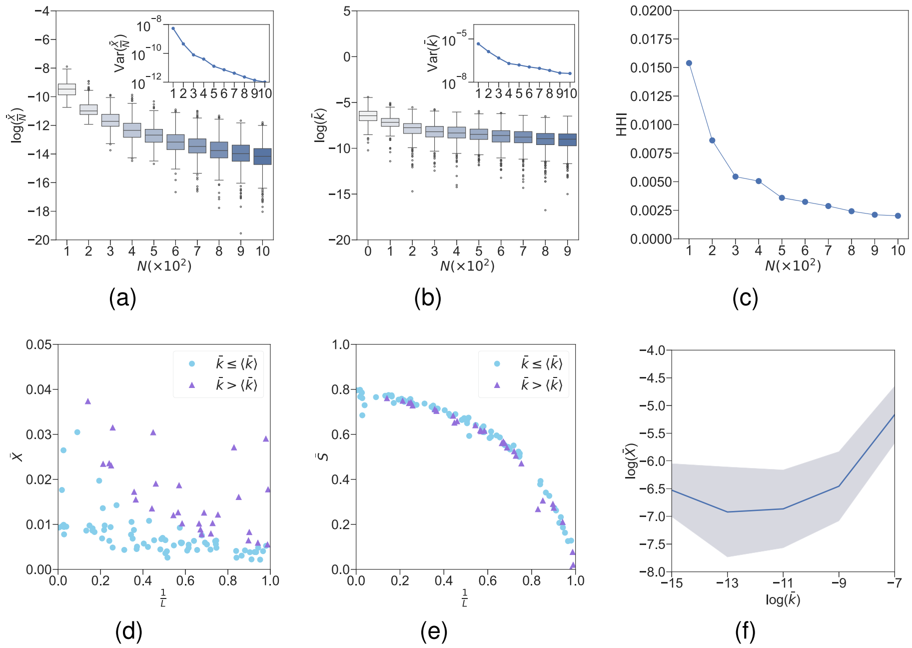

Fig. 3. Model prediction for ER networks. (a) Box plots for the distribution of steady-state market share ($\bar{X}$) for the unweighted capacity networks $\mathcal{G}_c^{(\mathrm{unw})}$ with $N \in [100, 1,000]$ nodes that have the same mean degree as BA capacity networks generated with $N \in [100, 1,000]$ and $m = 7$, with variances in the inset. (b) Box plots for the distribution of weighted degrees of nodes ($\bar{k}$) in the steady-state flow networks $\mathcal{G}^{(t)}$ where the ER networks have the same average degree as BA networks generated with $N \in [100, 1,000]$ and $m = 7$, with variances in the inset. On each number, the central mark is the median, and the bottom and top edges are the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data that are not outliers, and outliers are plotted individually using dots. (c) Scatter plot of the HHI for ER networks with the same average degree as BA networks generated with $N \in [100, 1,000]$ and $m = 7$. (d)–(e) Scatter plots of (d) the inverse of performance $\left(\frac{1}{L} = [\frac{1}{L_1}, \ldots, \frac{1}{L_N}]^\top\right)$ and steady-state market share $\bar{X}$; and (e) the inverse of performance and steady-state consumers' satisfaction $\bar{S}$ for all providers in the system. The providers are classified by the average weighted degree $\langle \bar{k} \rangle$ of the steady-state provider network $\mathcal{G}^{(t)}$. In this case, $\mathcal{G}_c^{(\mathrm{unw})}$ is an ER network with the same degree as the BA network generated by $N = 100, m = 7$. (f) Log-scale error plot for the steady-state consumer fluxes $\bar{k}$ and steady-state market share $\bar{X}$ for networks with $N \in [100, 1,000]$ and $m \in [1, 10]$. The line and the shaded area represent the mean value and the standard deviation of the distributions of $\bar{X}$ organized over five equally spaced intervals.

Additionally, we manually obtain the geographical coordinates of the doctors' office (latitude and longitude in seconds) using Google Maps [46]. The data on the types of health insurance plans including Medicaid and Commercial plans accepted by the doctors are collected from the Individual Provider Network Data provided by the New York State Department of Health on the website https://pndslookup.health.ny.gov.

B. Geographic Locations of Patients

The population density data for each zip code in NYC is obtained from the ZIP Code Tabulation Areas (ZCTAs) Dataset, which is released by the U.S. Census Bureau [47]. It consists of zip codes, population density, and geographical coordinates of the centroid of 192 zip code-level areas in NYC. In our setup, the patients are uniform-randomly distributed within each zip code-level area based on the population density of the zip code area. The boundaries of the zip code area, including latitude and longitude, are determined from the ZCTAs dataset. In this manner, the geographic locations of all the patients (in seconds) are specified in the system.

C. Data Analysis

In Fig. 4, we show a map of the population density of NYC at a zip code level along with the locations of the PCPs marked on the map. The darker color means densely populated areas and the lighter color reflects sparsely populated areas. From Fig. 4, we find that PCPs are more likely to locate in a darker area, while these circles are located close to each other. In fact, PCPs are clustered in densely populated areas – as seen from the spatial autocorrelation [48] of population density around the doctors (Moran's I $= 0.736$, $p = 0.001$).

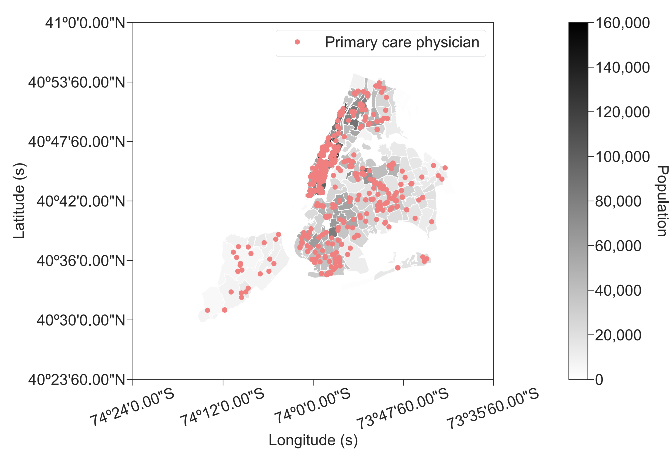

Fig. 4. Population map of NYC with red circles representing the PCPs in the database. The colored contours represent the population at zip code-level.

According to Hu et al. [39], customers who provide extreme (positive or negative) ratings on a product or a service are more inclined to express their opinion than those with moderate views, leading to a J-shaped distribution. Such a distribution is also found in our dataset of 9,508 ratings, where 77% of ratings are 5-star, 16% are 1-star, and only 7% in between. With an equivalent 5-point scale, Hu et al. [39] proposed that satisfaction is associated with a 4- or 5-star rating and classified ratings of 4 or higher as positive and those below as negative. We use the same classification, for all PCPs with more than five reviews on the website; those with less were discarded from the analysis.

D. Simulations

Based on the locations of patients and PCPs, we first set a distance threshold for each patient and allocate doctors within that distance to be accessible doctors to the patient. Then, we construct a capacity network $\mathcal{G}_c$ of doctors based on the accessibility. We choose distinct distance thresholds that determine provider accessibility. In quantifying accessibility, the standard approach involves gauging the distance to the nearest medical providers, a methodology substantiated by Guagliardo in their spatial analysis [22]. This entails patients exclusively selecting doctors located within a specific range from their domicile. Considering doctor P_i working at location (x_i, y_i) (latitude and longitude respectively) and patient C_l residing at a location (x_l, y_l), we first convert the latitude and longitude from seconds to radians and then calculate the distance d_{il} between them using the Haversine formula [49],

$$d_{il} = 2\arcsin\sqrt{\sin^2\left(\frac{x_i - x_l}{2}\right) + \cos x_i \cos x_l \sin^2\left(\frac{y_i - y_l}{2}\right)}. \tag{14}$$

For simplicity, we assume that each patient can access a maximum of two doctors in this model. Such an assumption is operationalized by randomly assigning, from the pool of doctors, a maximum of two within a circular area of radius d centered at the patient's location.

To obtain the initial number of positive and negative online ratings ($S_{i,+}^{(0)}$ and $S_{i,-}^{(0)}$) and that of textual reviews ($T^{(0)}$), we classify the online ratings of the PCPs in the dataset as positive if they are 4 or above and the corresponding textual reviews are also classified as positive. Then, we determine the performance of PCPs (L) based on the proportion of the positive textual reviews and compute the initial online satisfaction rates $S^{(0)}$ using (6). Similarly, using (9) and the inverse transform sampling method, each patient is assigned to a PCP based on the set of accessible PCPs and their initial online ratings and reviews from the database ($S^{(0)}$ and $T^{(0)}$). In this manner, the initial flow network of doctors $\mathcal{G}^{(0)}$ with the initial fraction of patients with the PCPs ($X^{(0)}$) are obtained. Numerically propagating (1), we run model simulations for 3,000 time units to ensure that the fraction of patients assigned to each doctor reaches equilibrium. The simulation parameters are the same as those introduced in Section III.

We examine simulation results at the steady state for different doctor networks in which the accessibility of PCPs is determined by various distance thresholds $d \in [1, 10]$ (km). Steady-state results for the market share of PCPs in Fig. 5(a) demonstrate that when the accessibility criterion relies on a larger d, the market share of PCPs tends to be more concentrated around the median over various doctor networks. This implies a reduction in the variability of the PCPs' success, where a larger proportion of patients can access PCP offices located at greater distances from their residential addresses. To investigate the competition among the providers, we also study the steady-state patient fluxes $\bar{k}$ as a function of d. The median value of $\bar{k}$ increases with increasing d, indicating that PCPs have more competitors in networks with larger d. Fig. 5(a) and (b) indicate that more patients can switch between the connected PCPs in networks with larger d, since the PCPs have more neighboring competitors in these networks. The HHI index decreases slowly with increasing d representing the increasing competition among the PCPs, see Fig. 5(c). As a result, PCPs could hardly succeed in attracting and retaining patients if located in a highly competitive market.

Further, we study the relationship among the performance of doctors L, online satisfaction rate $\bar{S}$, and the fraction of patients $\bar{X}$ who are with the doctor at the steady state, as shown in Fig. 5(d) and (e). Simulation results reveal that doctors with relatively high performance could attract more patients in Fig. 5(d), and potentially obtain high satisfaction rates in Fig. 5(e). Predictably, the high performance of PCPs is associated with more positive feedback from the patients. To investigate the competition experienced by PCPs, we classify PCP, based on $\bar{k}$, with respect to the average value, $\langle\bar{k}\rangle$. Similar to the synthetic networks, PCPs with larger $\bar{k}$ could attract more patients, see Fig. 5(d). The relationship between $\bar{X}$ and $\bar{k}$, is detailed in Fig. 5(f). PCPs experiencing higher $\bar{k}$ tend to achieve a higher $\bar{X}$ by competing with a large number of neighboring competitors, thereby taking the risk of gaining or losing their patients. Compared to the BA and ER networks (Figs. 2(f) and 3(f)), the PCP network has a less steep growth rate of the market share with respect to their degree.

Causal discovery helps us pinpoint the cause-and-effect relationships among these variables. We implement the PC algorithm proposed by Peter and Clark [50] on the simulated data, to determine the causal structure of the system variables, which can be represented by Directed Acyclic Graphs (DAGs). We start with a complete undirected graph: through iterative conditional independence tests, we remove edges and construct a skeleton of a DAG. To orient the edges of the skeleton into the final DAG, we apply deterministic rules that include orienting colliders, chains, and other specific configurations. The statistical test adopted for the PC algorithm is the Fisher's z score [51], and the statistical significance level is set to 0.05. Using the correlation among the steady-state values $\bar{k}$, $\bar{X}$, $\bar{S}$, and $\bar{Q}$ for various doctor networks, we discover statistically significant causal relationships among these variables, represented by a DAG in Fig. 6. This casual structure remains consistent across various doctor networks and different permutations of the input variables (since the PC algorithm is not order-invariant [52]). The DAG indicates that the market share $\bar{X}$ and the patients' satisfaction $\bar{S}$ can affect

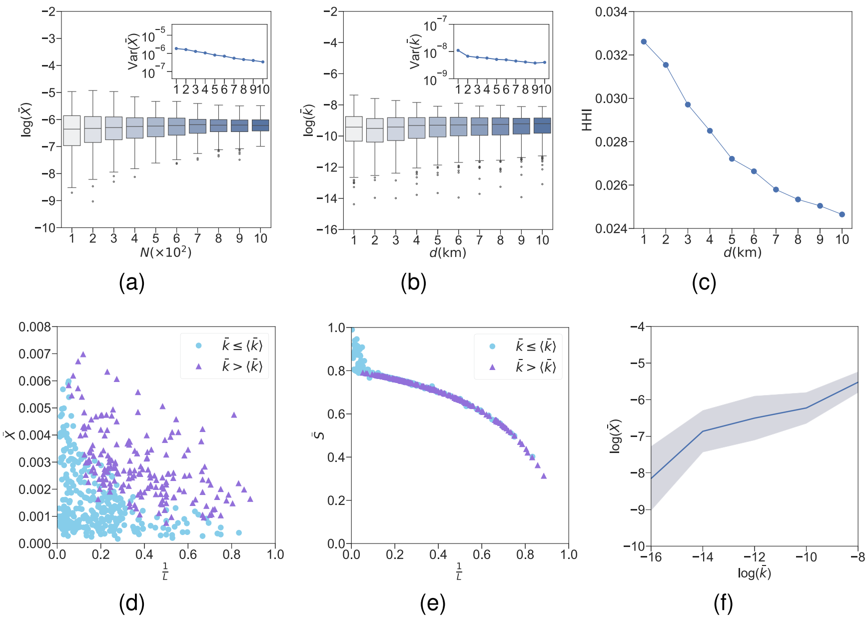

Fig. 5. Model prediction for doctor networks. (a) Box plots of the distribution of the fraction of patient $\bar{X}$ at the steady state over the flow networks where the PCP's accessibility is determined by the distance threshold $d \in [1, 10]$(km), with variances in the inset. (b) Box plots of the distribution of the steady-state fraction of patient fluxes ($\bar{k}$) over the flow networks in which accessibility of PCPs is determined by the distance threshold $d \in [1, 10]$(km), with variances in the inset. On each number, the central mark is the median, and the bottom and top edges are the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data that are not outliers, and outliers are plotted individually using dots. (c) Scatter plot of the HHI for doctor networks with different d. (d)-(e) Scatter plots of (d) the inverse of performance $\left(\frac{1}{L} = [\frac{1}{L_1}, \ldots, \frac{1}{L_N}]^\top\right)$ and the steady-state fraction of patients ($\bar{X}$); and (e) the inverse of performance and steady-state online satisfaction rate ($\bar{S}$) for all the PCPs in the system. The mean degree $\langle \bar{k} \rangle$ of the steady-state flow network is used to classify the PCPs. In this case, the accessibility of PCPs is determined by the distance threshold $d = 10$ (km). (f) Log-scale error plot for the steady-state patient fluxes $\bar{k}$ and steady-state market share $\bar{X}$ for the doctor network in which the PCP's accessibility is determined by the distance threshold $d \in [1, 10]$(km). The line and the shaded area represent the mean value and the standard deviation of the distributions of $\bar{X}$ organized over five equally spaced intervals.

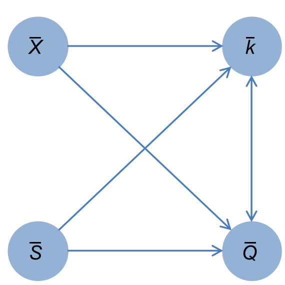

Fig. 6. Causal structure of the variables, including $\bar{k}$, $\bar{X}$, $\bar{S}$, and $\bar{Q}$ derived from the generated doctor networks where accessibility of PCPs is determined by distance threshold $d \in [1, 10]$ (km), where the significance level of edges is 0.05. Here, $\bar{k}$ is the steady-state patient fluxes, and $\bar{X}$, $\bar{S}$, and $\bar{Q}$ respectively represent the fraction of patients, patients' satisfaction, and healthcare quality for all the providers, across the various doctor networks. All the reported edges have a p value less than 0.05 (largest p value of 0.009 for the edge from $\bar{X}$ to $\bar{Q}$). None of the non-reported edges are close to the significance threshold (lowest p value of 0.253 for the edge from $\bar{X}$ to $\bar{S}$).

the patient fluxes $\bar{k}$ through the PCPs and the quality of care $\bar{Q}$. The causal link towards $\bar{k}$ can be attributed to the fact that the number of patients that a PCP cares for affects the extent to which they will gain or lose patients and the overall practice quality. Simultaneously, patients' satisfaction rate from the ORS plays a significant role in influencing their choice of a PCP. However, the orientation of the causal relationship between patient fluxes $\bar{k}$ and quality of care $\bar{Q}$ is difficult to determine from the existing dataset.

V. Conclusion

We propose a network-based dynamical model to study consumers' decision-making processes when selecting providers based on online reviews. The model utilizes two network structures: one to describe the accessibility of providers to the consumers and the other to track the consumer fluxes between providers. The former (capacity network) is static, whereby we hypothesize that each consumer has access to at most two providers, assigned to them once and for all. The latter (flow network) is, instead, time-varying to describe the intricate process by which providers compete for consumers, who may decide to change a provider for another on the basis of their own experience and the information they gather through the ORS. While capturing many salient features of decision-making processes, the proposed model is parsimonious in nature. We formulate the model as a discrete-time Markov process that is amenable to some analytical treatment. We can prove the well-posedness of the model, in terms of the range of variations of the number of consumers subscribed to each provider as a function of the provider accessibility.

We present simulation results for a number of case studies in which we vary the size of the market and the topology of the capacity network to simulate different scenarios of accessibility of providers to consumers. We specifically consider two canonical network models for the capacity network: Barabási-Albert (BA) and Erdaős-Rényi (ER). In a BA network, the network degree tends to follow a power-law distribution, with several nodes having a degree much larger than the average and connecting dense sub-networks of low-degree nodes. Hence, there are providers who are accessible to a large portion of the consumers and can compete for consumers with many other providers. These hubs coexist with a majority of providers whose accessibility is limited to a small group of consumers, mostly competing with providers who have the same reach. In an ER network, such a hierarchical structure is not present, whereby providers are practically equivalent with respect to their accessibility to consumers. We find that irrespective of the network type, as the market size increases (larger number of providers), the level of competition among providers diminishes, so that consumers tend to remain with their providers rather than switching to others. The consumers' satisfaction, together with the market share, impacts consumer fluxes between providers and the overall quality of service experienced by the consumers.

As a real-world test case, we apply our model to a system of PCP selection by patients in NYC based on online ratings and geographical proximity. For each physician in the dataset, the office location, online ratings, and textual reviews are important to distinguish the characteristics of decision-making. We observe a majority of PCPs being clustered together, generally in the densely populated areas of the city. Similar to our findings in the canonical networks, in this network, the satisfaction of patients from the ORS, combined with the market share, significantly influences the gain or loss of patients and the healthcare quality experienced by the patients. The study of the causal relationships between the patient fluxes between PCPs, quality of care, patients' satisfaction, and market share at the steady state reveals that the number of patients that a PCP provides care to can result in gaining or losing patients and affect the quality of their care. Patients' satisfaction rate from the ORS also plays a significant role in influencing their choice of PCP.

Our study is not free of limitations. First, our modeling approach is based on the simplifying assumption that a consumer can access a maximum of two providers. This simplifying assumption was necessary to formulate the model in terms of classical graph theory, rather than resorting to hypergraphs to capture selection among more than two providers. Future work shall seek to expand the model to hypergraphs to overcome this limitation. Second, our analytical insight is presently limited to general claims about the well-posedness of the model, whereby we rely on computer simulations to garner insight into the factors that shape decision-making. Future work could attempt further analytical research, potentially exploring linearized model formulations around the steady-state. For example, one could explore sensitivity analysis with respect to model parameters, similar to those studied by De Lellis et al. [53], to detail the relationship between local topological features of the capacity network and the overall response of the system. Third, we acknowledge that the analysis of the real-world problem of accessibility of providers is preliminary and requires additional work before supporting robust conclusions for the medical sector. In particular, we warn prudence against the definition of accessibility based on only physical proximity; transportation and costs are, for example, factors that should be considered when defining accessibility.

Acknowledgment

The authors thank Yindalon Aphinyanaphongs, Oded Nov, and Rayan Succar for useful discussions. *Dataset and Code Availability*: All data and codes that support the findings of this study are openly available on https://github.com/dynamicalsystemslaboratory/Consumer_Provider_System.

References

[1] P. Chatterjee, "Online reviews: Do consumers use them?," *NA - Adv. Consum. Res.*, vol. 28, pp. 129–133, 2001.
[2] C. M. Cheung, M. K. Lee, and N. Rabjohn, "The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities," *Internet Res.*, vol. 18, no. 3, pp. 229–247, 2008.
[3] A. M. Alkalbani, A. M. Ghamry, F. K. Hussain, and O. K. Hussain, "Predicting the sentiment of SaaS online reviews using supervised machine learning techniques," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2016, pp. 1547–1553.
[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2004, pp. 168–177.
[5] R. Filieri and F. McLeay, "E-wom and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews," *J. Travel Res.*, vol. 53, no. 1, pp. 44–57, 2014.
[6] C. M. Cheung and M. K. Lee, "What drives consumers to spread electronic word of mouth in online consumer-opinion platforms," *Decis. Support Syst.*, vol. 53, no. 1, pp. 218–225, 2012.
[7] R. Janakiraman and R. Niraj, "The impact of geographic proximity on what to buy, how to buy, and where to buy: Evidence from high-tech durable goods market," *Decis. Sci.*, vol. 42, no. 4, pp. 889–919, 2011.
[8] P. G. Patterson, L. W. Johnson, and R. A. Spreng, "Modeling the determinants of customer satisfaction for business-to-business professional services," *J. Acad. Marketing Sci.*, vol. 25, no. 1, pp. 4–17, 1996.
[9] G. H. McDougall and T. Levesque, "Customer satisfaction with services: Putting perceived value into the equation," *J. Serv. Marketing*, vol. 14, no. 5, pp. 392–410, 2000.
[10] J. Tang, X. Hu, and H. Liu, "Social recommendation: A review," *Social Netw. Anal. Mining*, vol. 3, pp. 1113–1133, 2013.
[11] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora, "Distance matters: Geo-social metrics for online social networks," in *Proc. 3rd Conf. Online Social Netw.*, 2010, p8.
[12] H. Gao, J. Tang, and H. Liu, "gSCorr: Modeling geo-social correlations for new check-ins on location-based social networks," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, 2012, pp. 1582–1586.
[13] M. J. Osborne and C. Pitchik, "Equilibrium in hotelling's model of spatial competition," *Econometrica: J. Econometric Soc.*, vol. 55, pp. 911–922, 1987.
[14] M. E. Porter et al., "Clusters and the new economics of competition," *Harvard Bus. Rev.*, vol. 76, no. 6, pp. 77–90, 1998.
[15] M. E. Porter, "The five competitive forces that shape strategy," *Harvard Bus. Rev.*, vol. 86, no. 1, 2008, Art. no. 78.
[16] W. Quattrociocchi, G. Caldarelli, and A. Scala, "Opinion dynamics on interacting networks: Media competition and social influence," *Sci. Rep.*, vol. 4, no. 1, pp. 1–7, 2014.
[17] C. G. Antonopoulos and Y. Shang, "Opinion formation in multiplex networks with general initial distributions," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 2852.
[18] N. Hudson and H. Khamfroush, "Behavioral information diffusion for opinion maximization in online social networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1259–1268, Apr.–Jun. 2021.
[19] N. E. Friedkin, A. V. Proskurnikov, R. Tempo, and S. E. Parsegov, "Network science on belief system dynamics under logic constraints," *Sci.*, vol. 354, no. 6310, pp. 321–326, 2016.
[20] J. Hausman and D. McFadden, "Specification tests for the multinomial logit model," *Econometrica: J. Econometric Soc.*, vol. 52, pp. 1219–1240, 1984.
[21] J. H. Kim, F. Pagliara, and J. Preston, "The intention to move and residential location choice behaviour," *Urban Stud.*, vol. 42, no. 9, pp. 1621–1636, 2005.

[22] M. F. Guagliardo, "Spatial accessibility of primary care: Concepts, methods and challenges," *Int. J. Health Geographics*, vol. 3, no. 1, pp. 1–13, 2004.
[23] J. R. Sharkey, "Measuring potential access to food stores and food-service places in rural areas in the us," *Amer. J. Prev. Med.*, vol. 36, no. 4, pp. S151–S155, 2009.
[24] M. Kanagawa and T. Nakata, "Assessment of access to electricity and the socio-economic impacts in rural areas of developing countries," *Energy Policy*, vol. 36, no. 6, pp. 2016–2029, 2008.
[25] K. E. Smoyer-Tomic et al., "The association between neighborhood socioeconomic status and exposure to supermarkets and fast food outlets," *Health Place*, vol. 14, no. 4, pp. 740–754, 2008.
[26] M. Delasay, A. Ingolfsson, B. Kolfal, and K. Schultz, "Load effect on service times," *Eur. J. Oper. Res.*, vol. 279, no. 3, pp. 673–686, 2019.
[27] E. Reinhart, "Doctors aren't burned out from overwork. we're demoralized by our health system," *NY Times*, vol. 5, p. 18, 2023.
[28] M. E. Porter and E. O. Teisberg, *Redefining Health Care: Creating Value-Based Competition on Results*. Boston, MA, USA: Harvard Bus. Press, 2006.
[29] D. C. Jaravaza and P. Chitando, "The role of store location in influencing customers' store choice," *J. Emerg. Trends Econ. Manage. Sci.*, vol. 4, no. 3, pp. 302–307, 2013.
[30] J. R. Hanaysha, M. E. Al Shaikh, and H. M. Alzoubi, "Importance of marketing mix elements in determining consumer purchase decision in the retail market," *Int. J. Serv. Sci. Manage. Eng. Technol.*, vol. 12, no. 6, pp. 56–72, 2021.
[31] D. Rigby, "The future of shopping," *Harvard Bus. Rev.*, vol. 89, no. 12, pp. 65–76, 2011.
[32] M. Söderlund, "Customer satisfaction and its consequences on customer behaviour revisited: The impact of different levels of satisfaction on word-of-mouth, feedback to the supplier and loyalty," *Int. J. Serv. Ind. Manage.*, vol. 9, no. 2, pp. 169–188, 1998.
[33] B. A. Sparks and V. Browning, "Complaining in cyberspace: The motives and forms of hotel guests' complaints online," *J. Hospitality Marketing Manage.*, vol. 19, no. 7, pp. 797–818, 2010.
[34] W. W. Moe and M. Trusov, "The value of social dynamics in online product ratings forums," *J. Marketing Res.*, vol. 48, no. 3, pp. 444–456, 2011.
[35] S. Boccaletti et al., "The structure and dynamics of networks with higher order interactions," *Phys. Rep.*, vol. 1018, pp. 1–64, 2023.
[36] J. Heikkilä, "From supply to demand chain management: Efficiency and customer satisfaction," *J. Operations Manage.*, vol. 20, no. 6, pp. 747–767, 2002.
[37] H. Liao and A. Chuang, "A multilevel investigation of factors influencing employee service performance and customer outcomes," *Acad. Manage. J.*, vol. 47, no. 1, pp. 41–58, 2004.
[38] D. Gavilan, M. Avello, and G. Martinez-Navarro, "The influence of online ratings and reviews on hotel booking consideration," *Tourism Manage.*, vol. 66, pp. 53–61, 2018.
[39] N. Hu, J. Zhang, and P. A. Pavlou, "Overcoming the j-shaped distribution of product reviews," *Commun. ACM*, vol. 52, no. 10, pp. 144–147, 2009.
[40] A. M. Law, W. D. Kelton, and W. D. Kelton, *Simulation Modeling and Analysis*. New York, NY, USA: Mcgraw-hill, 2007.
[41] R. D. Buzzell, B. T. Gale, and R. G. Sultan, "Market share-a key to profitability," *Harvard Bus. Rev.*, vol. 53, no. 1, pp. 97–106, 1975.
[42] A. N. Berger, L. F. Klapper, and R. Turk-Ariss, "Bank competition and financial stability," in *Handbook of Competition in Banking and Finance*. Cheltenham, U.K.: Edward Elgar Publishing, 2017, pp. 185–204.
[43] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Modern Phys.*, vol. 74, no. 1, 2002, Art. no. 47.
[44] A. Vázquez, "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations," *Phys. Rev. E*, vol. 67, no. 5, 2003, Art. no. 056104.
[45] L. Richardson, "Beautiful soup documentation," 2007. Accessed: Nov. 5, 2022. [Online]. Available: https://www.crummy.com/software/BeautifulSoup/bs4/doc/
[46] Google Maps, Accessed: Mar. 14, 2023. [Online]. Available: https://www.google.com/maps/place/New+York+NY
[47] U. C. B., "Zip code tabulation areas (ZCTAs) are generalized areal representations of united states postal service (USPS) zip code service areas," 2011. [Online]. Available: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html
[48] Y. Chen, "New approaches for calculating Moran's index of spatial autocorrelation," *PLoS ONE*, vol. 8, no. 7, 2013, Art. no. e68336.
[49] K. Gade, "A non-singular horizontal position representation," *J. Navigation*, vol. 63, no. 3, pp. 395–417, 2010.
[50] P. Spirtes, C. N. Glymour, and R. Scheines, *Causation, Prediction, and Search*. Cambridge, MA, USA: MIT Press, 2000.
[51] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.
[52] D. Colombo et al., "Order-independent constraint-based causal structure learning," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3741–3782, 2014.
[53] P. De Lellis, M. Ruiz Marín, and M. Porfiri, "Modeling human migration under environmental change: A case study of the effect of sea level rise in Bangladesh," *Earth's Future*, vol. 9, no. 42021, Art. no. e2020EF001931.

Tian Gan received the B.Sc. degree in communications engineering from Northeast Forestry University, Harbin, China, in 2018 and the M.Sc. degree in electrical engineering from Delft University of Technology, Delft, Netherlands, in 2021. She is currently working toward the Ph.D. degree with New York University Tandon School of Engineering, Brooklyn, NY, USA. Her research interests include network science, data science, and complex systems.

Rishita Das received the B.Tech. and M.Tech. degrees in aerospace engineering from the Indian Institute of Technology Kharagpur, Kharagpur, India, in 2016 and the Ph.D. degree in aerospace engineering from Texas A&M University, College Station, TX, USA, in 2021. She is currently an Assistant Professor with the Department of Aerospace Engineering, the Indian Institute of Science, Bangalore, India. She was a Postdoctoral Associate in mechanical and aerospace engineering with New York University, Brooklyn, NY, USA. Her research interests include dynamical systems theory, data science, and complex systems. She is an Amelia Earhart Fellow and a DAAD WISE Scholar.

Maurizio Porfiri (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in engineering mechanics from Virginia Tech, Blacksburg, VA, USA, in 2000 and 2006, respectively and the "Laurea" (Hons.) degree in electrical engineering and the Ph.D. degree in theoretical and applied mechanics from the Sapienza University of Rome, Rome, Italy and the University of Toulon, Toulon, France (under a dual-degree program between Italy and France), in 2001 and 2005, respectively. He is currently an Institute Professor with the New York University Tandon School of Engineering, Brooklyn, NY, USA, having appointments with the Center for Urban Science and Progress, Department of Mechanical and Aerospace Engineering, Department of Biomedical Engineering, and Department of Civil and Urban Engineering. He is the Director of the Center for Urban Science and Progress and the Dynamical Systems Laboratory. His research interests in complex systems span applications from mechanics to behavior, public health, and robotics. He has authored more than 400 journal publications. He has been included in the "Brilliant 10" list of Popular Science in 2010. His research was featured in major media outlets, including CNN, NPR, Scientific American, and Discovery Channel. Dr. Porfiri was the recipient of the National Science Foundation CAREER Award, the Outstanding Young Alumnus Award by the College of Engineering, Virginia Tech, the American Society of Mechanical Engineers (ASME) Gary Anderson Early Achievement Award, the ASME DSCD Young Investigator Award, the ASME C.D. Mote, Jr. Early Career Award, and the Research Excellence Award from New York University Tandon School of Engineering. His other significant recognitions include invitations to the Frontiers of Engineering Symposium and the Japan-America Frontiers of Engineering Symposium organized by the National Academy of Engineering and to the third and fourth World Laureate Forums. He has served on the Editorial Board of the *ASME Journal of Dynamics Systems, Measurements and Control*, *the ASME Journal of Vibrations and Acoustics*, *Flow: Applications of Fluid Mechanics*, the IEEE Control Systems Letters, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: REGULAR PAPERS, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, *Mathematics in Engineering*, and *Mechatronics*. He is a Fellow of the ASME.