Chapter 6
Towards Identifying and Predicting Spatial Epidemics on Complex Meta-population Networks

Xiang Li, Jian-Bo Wang, and Cong Li

Abstract In the past decade, the network science community has witnessed huge advances in the threshold theory, prediction and control of epidemic dynamics on complex networks. While along with the understanding of spatial epidemics on meta-population networks achieved so far, more challenges have opened the door to identify, retrospect, and predict the epidemic invasion process. This chapter reviews the recent progress towards identifying susceptible-infected compartment parameters and spatial invasion pathways on a meta-population network as well as the minimal case of two-subpopulation version, which may also extend to the prediction of spatial epidemics as well. The artificial and empirical meta-population networks verify the effectiveness of our proposed solutions to the concerned problems. Finally, the whole chapter concludes with the outlook of future research.

6.1 Introduction

After around 70 years of the seminal work of Norbert Wiener "Cybernetics: or the Control and Communication in the Animal and the Machine" [1], Wiener's great thinking still presents fundamental impacts to many folds of the human society in the era of networking world and Big Data today, ranging from modelling and

A part of this chapter has been contributed to the 2016 IEEE Conference on Norbert Wiener in the twenty-first century (21CW) to memorize Dr. Norbert Wiener. This work was partly supported by the Natural Science Fund for Distinguished Young Scholar of China (No. 61425019), the National Natural Science Foundation (Nos. 61273223, 61603097), the Key Project of National Social Science Fund (No. 12&ZD218) of China, Shanghai SMEC-EDF Shuguang Project (No. 14SG03), and Natural Science Foundation of Shanghai (No. 16ZR1446400).

X. Li (✉) • J.-B. Wang • C. Li
Adaptive Networks and Control Lab, Department of Electronic Engineering and Research Center of Smart Networks & Systems, School of Information Science & Engineering, Fudan University, 200433, Shanghai, People's Republic of China
e-mail: lix@fudan.edu.cn

N. Masuda, P. Holme (eds.), *Temporal Network Epidemiology*,
Theoretical Biology, DOI 10.1007/978-981-10-5287-3_6

feedback-loop analysis to stability and control of categories of systems and subjects, whatever large-scale or simply structured, linear or nonlinear, low dimensional or extremely high dimensional. The communications among humans and machines in the eyes of Norbert Wiener in 1950s were generally assumed as point-to-point or neglected as regularly structured in the scope of classic graph theory [2, 3]. Afterwards, Erdős and Rényi extended the graph description with uncertainty and randomness, and proposed the random graph theory in 1960s [4]. In the following decades the flourishing information and communication techniques have pushed the whole human society to a networking village of today, while the understanding of dominant yet hidden connectivity patterns of the communications among humans and machines were not revisited until recently.

The discovery of small-world and scale-free features in 1998–1999 has been verified in ubiquitous complex networks [5, 6], which have attracted the world-wide attention to the new emergence of network science. The popular concerns cover not only the topological complexity of a large-scale complex network system but also the interdependence between the infrastructure and the collective performance of such networks [7–10]. Typically, from the viewpoint of system and control, the precise mathematic description and appropriate models of a complex network play a significant role to achieve the desirable performance in return. However, in the situations of large-scale spatial prevalence of diseases in human populations, for example, such a solution may be infeasible if the availability of accurate data collections is far from sufficiently satisfactory.

Nevertheless the global outbreaks of prevalent infectious diseases in recent decades have led to great social, economic, and public health loss [11–14], which is partially due to the urbanization process and, in particular, the wide-establishment of long-distance public transportation networks (e.g., world-wide air-line web) and urban public commuting systems (e.g., subway and metro networks) to facilitate the dissemination of pathogens accompanied with passengers [15, 16]. Academia has witnessed that prediction and control of epidemic dynamics in networks as a flourishing research topic with interdisciplinary approaches [17–20]. However, more challenging problems arising from the epidemic prevalence on a meta-population network have not received adequate attentions, such as identifying the parameters of epidemic network systems and the epidemic invasion pathways on a meta-population network, which, ignored previously, certainly play important roles in evaluating the intensity of outbreak of epidemics among human patches/populations.

Assume the seed of a disease/virus as the input signal to the whole human population system, and the observed patient samples as the system output. Then, the spatial invasion of the disease inside the human population is obscure as a black box to be identified, and this system combines many factors such as human mobility patterns (commuting and long-distance traveling) and mathematical epidemiology as well. Therefore, identifying such an epidemic process with the interplay of complex networks and the human population is a challenge to public health-care administrative agency when predicting the large-scale spatial prevalence of a disease and announcing counter strategies.

The theory of system identification has been used to estimate the epidemic parameters of a complex system which are described by ordinary differential equations (ODE) such as HIV/AIDS epidemic dynamics [21]. Another related topic is inferring network topology by utilizing the information about a dynamics process on networks [22, 23]. Note that system identification and network inference techniques are not fit to handle the epidemic process on meta-population networks which are stochastic, high-dimensional, and multi-scale. Besides, source identification on complex networks is a close and popular topic. Some source identification algorithms [24, 25] have been designed for information/contact networks, but they are not feasible in identifying the invasion processes on meta-population networks.

Many instructive methods have also been proposed to explore the spatial spread of an epidemic process on meta-population networks. Maeno [26] inferred the epidemic network between eleven countries and areas during SARS in 2003 by analysing the epidemic time series. Reference [27] extracted the most likely epidemic transmission trees of the 1918 influenza pandemic in England, Wales and the United States. Some methods based on machine learning were also proposed to infer the epidemic networks from surveillance data [28–30]. Gautreau et al. presented a measure of the average arrival time to characterize the minimum-distance path from subpopulation i to subpopulation j over all possible paths [31], and the average arrival time-based shortest path tree is constructed by assembling all the shortest paths from the seed subpopulation to any other subpopulation in a networked meta-population. Balcan et al. proposed a Monte Carlo maximum likelihood method to produce a most likely infection tree [32]. They constructed the minimum spanning tree from the seed subpopulation to minimize the distance. Recently, Brockmann and Helbing [15] proposed a new concept called "effective distance" to predict the disease arrival time. From node/location i to node/location j, the effective distance D_{ij} is defined as the minimum sum of effective lengths over all reachable branches along this path. The set of shortest paths to all other nodes from seed node i constitutes a shortest path tree, illustrating the most probable paths from the root to other nodes. On the other hand, approaches based on machine learning such as genetic algorithm [28–30] has been used to extract epidemic transmission networks.

Note that some of the above works didn't distinguish epidemic transmission network and invasion pathways/trees. In fact, these two concepts are a bit different, and very few work has discussed the parameter identification of a meta-population network system. Here a natural problem poses itself that whether the parameters and epidemic invasion process can be identified from the infection data of populations and network topology? To get a better understanding of how the contagion diffuses via an invasion process on network, more topics deserve further efforts: (i) So far, there are few works on identification of parameters of a meta-population network which an epidemic is occurring on. New questions such as the following ones are raised: How to use the data from the limited epidemic realizations to infer the system parameters as accurate as possible? Does a more appropriate model of individual mobility exist? (ii) Identification of spatial invasion pathways is to uncover the channels by which the hosts transmit viruses in a spatially structured population with the infection data. In a large-scale meta-population network, the complex pattern of pathways challenges the methodology to identify the epidemic invasion pathways in a meta-population network.

In this chapter, we review our series of work in recent years [33–37] on identifying parameters of the susceptible-infected model and spatial invasion pathways on a meta-population network as well as the minimal case of two-subpopulation version, which may also extend to the prediction of spatial epidemics as well. The remainder of this chapter is arranged as follows. Section 6.2 gives the detailed description of preliminaries. Section 6.3 introduces the parameter identification of epidemic models on a meta-population network. Section 6.4 contributes the inference of epidemic invasion pathways in a meta-population network with both methodologically and example verifications. In Sect. 6.5, extending the steps of the previous sections, the prediction of spatial epidemic transmission comes with several feasible methods. Finally, Sect. 6.6 concludes the whole chapter with outlook in future research.

6.2 Preliminary

A meta-population network, which was originated from the meta-population model proposed by Richard Levins [38] to explore spatial ecology, embeds public transportation networking systems to model and uncover nontrivial patterns of spatial prevalence of global infectious diseases in the past years [15, 31, 32, 39, 40]. In this section, we introduce the meta-population network model and the susceptible-infected (SI) compartment epidemic dynamic as well. In this chapter, we consider the discrete-time dynamics.

6.2.1 The Compartment Model with SI Reaction Dynamics

The well-known susceptible-infected (SI) compartment model (Fig. 6.1), which is the simplest version in the epidemic compartment family, generally describes the early stage of prevalence of viruses/pathogen [24, 25, 41], especially in the situation of non-recovery. In such a population, the states of individuals are stratified into two compartments (classes): susceptible to the infection of the pathogen; and infected by the pathogen. Generally, we assume that all individuals are homogenously mixing in the population. The state transition of an individual between two compartments

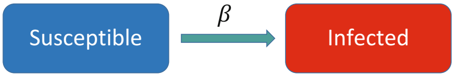

Fig. 6.1 Schematic illustration of the SI compartment model, where β denotes the infection rate

is governed by the following reaction process: When a susceptible individual meets (i.e., has the contacts) with an infected individual in a unit time, the susceptible individual will be infected with an infection rate β.

6.2.2 *The Two-Subpopulation Version of a Meta-population Network*

Before describing a general meta-population network, we first introduce a minimal meta-population network containing two subpopulations (labelled as 1 and 2 as shown in Fig. 6.2) with SI epidemic compartments. We assume the infection process evolves as a discrete-time system, and subpopulation 1 is infected initially (In this case of simulation, we assume 1 individual is infected among all 10,000 individuals in subpopulation 1). During each time step, the reaction takes place in each subpopulation if it contains two classes of individuals (susceptible and infected). Denote p_{12} (p_{21}) the diffusion rate of individuals transferring from subpopulation 1 to 2 (2 to 1), which are often not symmetric, i.e., $p_{12} \neq p_{21}$. Besides, an individual in subpopulation 1 (2) chooses jumping to subpopulation 2 (1) at diffusion rate p_{12} (p_{21}), i.e., the so-called diffusion process. Therefore, the probability an individual stays in subpopulation 1 (2) is $1 - p_{12}$ ($1 - p_{21}$).

Therefore, without considering the diffusion of new increment of infected individuals after reaction, the whole evolution dynamics is described as

$$\begin{cases} \langle I_1(t+1) - I_1(t) \rangle = \beta I_1(t)\dfrac{S_1(t)}{N_1(t)} + p_{21}I_2(t) - p_{12}I_1(t), \\ \\ \langle I_2(t+1) - I_2(t) \rangle = \beta I_2(t)\dfrac{S_2(t)}{N_2(t)} + p_{12}I_1(t) - p_{21}I_2(t), \end{cases} \tag{6.1}$$

where $\langle \cdot \rangle$ represents the expectation of the corresponding terms, $N_1(t)$ ($N_2(t)$) denotes the number of individuals in subpopulation 1 (2) at time t, $I_1(t)$ ($I_2(t)$) denotes the number of infected individuals in subpopulation 1 (2), $S_1(t)$ ($S_2(t)$)

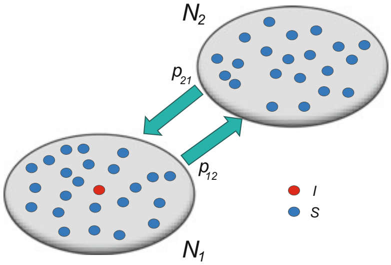

Fig. 6.2 Schematic representation of a minimal meta-population network with the SI model. At initial time, subpopulation 1 is infected (containing at least one infected individual (*red*)), and subpopulation 2 is susceptible (all are susceptible individuals (*blue*)) (From Wang et al. [33])

denotes the number of susceptible individuals in subpopulation 1 (2). The first term of the right-hand side (RHS) in Eq. (6.1) represents the new increment of infected individuals $\langle \Delta_R I_i(t) \rangle = \beta I_i(t) \frac{S_i(t)}{N_i(t)}, i = 1, 2$, after reaction from t to $t+1$. The second and third terms of RHS in Eq. (6.1) represent the diffusion of infected individuals in the diffusion process. As mentioned above, we do not consider the diffusion of new increment of infected individuals after reaction in this case. Besides, the evolution of susceptible individuals is similar with the infected individuals.

6.2.3 The General Description of a Meta-population Network

Extending the minimal version as two subpopulations to the general case of a meta-population network, we divide the whole population (generally, such a population covers a large-scale spatial region of a country or the whole world) into a number of subpopulations. In a meta-population network, a subpopulation is connected with others via a public transportation network, e.g., the air-line web, the high-way web to form the backbone of such a meta-population network. A subpopulation as a node in the network contains a number of individuals homogeneously mixed, and individuals travel between two subpopulations (nodes) via the public transportation means (edge) with some (fixed) diffusion rate. All edges are directed.

With the SI dynamics, the disease propagates in subpopulations and spreads among neighbouring subpopulations via the reaction-diffusion process in a unit time, as illustrated in Fig. 6.3. Denote N the number of subpopulations (nodes) of a meta-population network, and $N_i(t) = S_i(t) + I_i(t)$ is the population size of

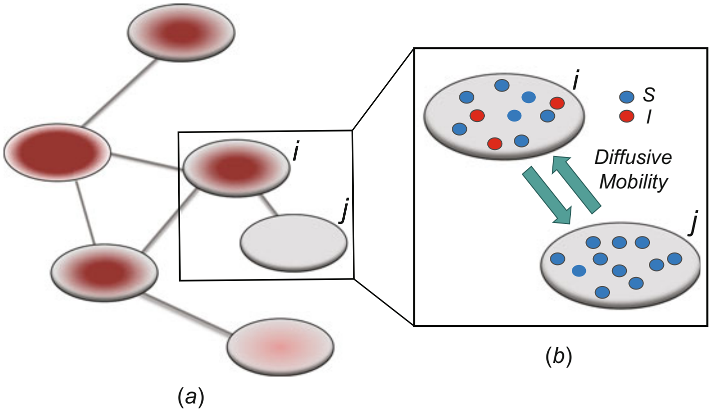

Fig. 6.3 Illustration of a networked meta-population model, which comprises six subpopulations that are coupled by the mobility of individuals. In each subpopulation, each individual can be in one of the two states (susceptible, infected), as shown in different colours. *Grey ones* represent susceptible subpopulations. *Red ones* represent infected subpopulations. *Light red* subpopulations denote less number of infected individuals than the *dark red ones*. Each individual can travel between the connected subpopulations. (**a**) A networked meta-population. (**b**) Two subpopulations (From Wang et al. [35])

subpopulation i at time t, where $S_i(t)$ is the number of susceptible individuals, and $I_i(t)$ is the number of infected individuals of subpopulation i at time t, respectively. Therefore, the intra-population epidemic dynamics in subpopulation i is governed by the SI model. Per unit time, the risk of infection of a susceptible individual within subpopulation i is characterized by $\lambda_i(t) = \beta I_i(t)/N_i(t)$ during the reaction process. Denote the probability that an individual (S or I) of subpopulation i moves to its neighbouring subpopulation j as diffusion rate p_{ij}, which describes the inter-population mobility dynamics. The symbol of diffusion rate $0 \leq p_{ij} = \frac{\langle w_{ij} \rangle}{\langle N_i \rangle} < 1$, where w_{ij} is the number of individuals moving from subpopulation i to j per unit time $(0 \leq \langle w_{ij} \rangle < \langle N_i \rangle)$.

Therefore, if we do not consider the diffusion of new increment of infected individuals after the reaction process, the evolution of an infected subpopulation i is described as follows:

$$\langle I_i(t+1) - I_i(t) \rangle = \beta I_i(t) \frac{S_i(t)}{N_i(t)} + \sum_{j=1, j \neq i}^{N} p_{ji} I_j(t) - \sum_{j=1, j \neq i}^{N} p_{ij} I_i(t), \tag{6.2}$$

which is investigated in Sect. 6.4.

When we consider the diffusion of new increment of infected individuals after the reaction, the evolution is described by

$$\begin{aligned}\langle I_i(t+1) - I_i(t) \rangle =& \beta I_i(t) S_i(t)/N_i(t) \\ &+ \sum_{j \neq i} \left\{ p_{ji}[I_j(t) + \langle \Delta_R I_j(t) \rangle] - p_{ij}[I_i(t) + \langle \Delta_R I_i(t) \rangle] \right\},\end{aligned} \tag{6.3}$$

where $\Delta_R I_j(t)$ is the increment of $I_j(t)$ after the reaction from t to $t+1$. We give the extensive investigation of the dynamics given by Eq. (6.3) in Sect. 6.5.

We now discuss the individual *mobility operator* to handle the presence of stochasticity and independence of individual mobility, where the number of successful migration of individuals between adjacent subpopulations is quantified by a binomial or a multinomial process, respectively. If the focal subpopulation i only has one neighbouring subpopulation j, the number of individuals in a given compartment $\mathscr{X}$ ($\mathscr{X} \in \{S, I\}$ and $\sum_{\mathscr{X}} \mathscr{X}_i = N_i$) transferred from i to j per unit time, $\mathscr{T}_{ij}(\mathscr{X}_i)$, is generated from a binomial distribution with probability p_{ij} representing the diffusion rate and the number of trials $\mathscr{X}_i$, i.e.,

$$Binomial(\mathscr{T}_{ij}, \mathscr{X}_i, p_{ij}) = \frac{\mathscr{X}_i!}{\mathscr{T}_{ij}!(\mathscr{X}_i - \mathscr{T}_{ij})!} p_{ij}^{\mathscr{T}_{ij}} (1 - p_{ij})^{(\mathscr{X}_i - \mathscr{T}_{ij})}, \tag{6.4}$$

where $1 - p_{ij}$ denotes the probability of an individual staying in subpopulation i.

If the focal subpopulation i has multiple neighbouring subpopulations $j_1, j_2, \ldots, j_k$, with k representing i's degree, the numbers of individuals in a given compartment $\mathscr{X}$ moving from i to $j_1, j_2, \ldots, j_k$ are generated from a multinomial distribution with probabilities $p_{ij_1}, p_{ij_2}, \ldots, p_{ij_k}$ representing the diffusion rates on the edges emanated from subpopulation i and the number of trails $\mathscr{X}_i$, i.e.,

$$
\begin{aligned}
&Multinominal(\{\mathscr{T}_{ij_\ell}\}, \mathscr{X}_i, \{p_{ij_\ell}\}) \\
&\quad = \frac{\mathscr{X}_i!}{\prod_\ell \mathscr{T}_{ij_\ell}!(\mathscr{X}_i - \sum_\ell \mathscr{T}_{ij_\ell})!}(\prod_\ell p_{ij_\ell}^{\mathscr{T}_{ij_\ell}})(1 - \sum_\ell p_{ij_\ell})^{(\mathscr{X}_i - \sum_\ell \mathscr{T}_{ij_\ell})},
\end{aligned}
\tag{6.5}
$$

where integer $\ell \in [1, k]$, term $1 - \sum_\ell p_{ij_\ell}$ denotes the probability of an individual staying in subpopulation i.

6.3 Epidemic Parameter Identification

The epidemic parameters of a networked meta-population include the infection rate and diffusion rate, which play an important role in the SI dynamics, while the stochastic epidemic dynamics and the limit of available data make such an identification task more difficult. In this section, we review the method to identify both parameters for a two-subpopulation network and an estimation of infection rate for a general network version.

6.3.1 The Case of Two-Subpopulation Model

We first describe one realization of the invasion process evolving as follows. At the beginning, subpopulation 1 has been initialized with one infected individual in this case. When time evolves, the number of infected individuals $I_1(t)$ of subpopulation 1 increases due to the SI reaction dynamics in this subpopulation. The epidemic arrival time (EAT) is defined as the first arrival time of infected individuals from an infected subpopulation moving to a neighbouring susceptible subpopulation. To address the EAT, some infected individual(s) will move (diffuse) to subpopulation 2, which finally succeed in infecting subpopulation 2. Therefore, recording the infection data (the number of infected individuals in subpopulation i at time t, i.e., $I_i(t), i = 1, 2$) of each subpopulation as the available data, we need to identify the unknown infection rate β and diffusion rate p_{12}.

At the early stage of epidemic dynamics, we can approximate $S_i(t) \approx N_i(t), i = 1, 2$ ($I_i(0) \ll N_i(0)$) and therefore simplify Eq. (6.1) as

$$
\langle I_1(t+1) - I_1(t) \rangle + \langle I_2(t+1) - I_2(t) \rangle \approx \beta(I_1(t) + I_2(t)). \tag{6.6}
$$

Denote $I(t)$ the number of infected individuals in all subpopulations at time t, i.e., $I(t) = I_1(t) + I_2(t)$. Traditionally, the RHS of the above equation accounts for an exponential growth of the number of infected individuals, and β is regarded as the Malthusian growth rate. Thus, we rewrite Eq. (6.6) in the compact form as $I(t) \approx e^{\beta(t-0)}I(0)$. Considering $\ln[I(0)] \ll \ln[I(t)]$, $(0 \ll t)$, we have $\beta \sim \frac{\ln[I(t)]}{t}$. Therefore, we estimate the infection rate β by fitting the slope of $\ln[I(t)]$.

We now discuss how to identify diffusion rate p_{12}. Repeat the invasion of subpopulation 2 from subpopulation 1 until we record the epidemic arrival time to subpopulation 2, i.e., the disease/virus finally lands in subpopulation 2 and starts the local infection. We investigate the period from the initial time ($t = 0$) to the epidemic arrival time (t_{EAT}) that the first $\mathscr{H}$ individuals from subpopulation 1 invade subpopulation 2. From $t_{EAT} - 1$ to t_{EAT}, we get

$$\begin{cases} \langle I_1(t_{EAT}) - I_1(t_{EAT}-1)\rangle = \beta I_1(t_{EAT}-1) - p_{12}I_1(t_{EAT}-1), \\ \langle I_2(t_{EAT}) - I_2(t_{EAT}-1)\rangle = p_{12}I_1(t_{EAT}-1). \end{cases} \tag{6.7}$$

The likelihood function about the first $\mathscr{H}$ infected individuals from subpopulation 1 traveling to subpopulation 2 at time t_{EAT} is

$$\mathrm{P}(\mathscr{H}, t_{EAT}-1) = C^{\mathscr{H}}_{I_1(t_{EAT}-1)}(1-p_{12})^{I_1(t_{EAT}-1)-\mathscr{H}}p_{12}^{\mathscr{H}} \times \prod_{i=1}^{\eta-1}(1-p_{12})^{I_1(i)}, \tag{6.8}$$

where $t_{EAT} - 1 = \eta$, $\eta(\eta \geq 1)$ is an integer. If there are s ($s \geq 1$) rounds of repeated realizations of invasion processes, the joint likelihood function is given by

$$\begin{aligned} &\mathrm{P}(\mathscr{H}^{\{1\}}, t_1; \mathscr{H}^{\{2\}}, t_2; \cdots; \mathscr{H}^{\{s\}}, t_s) \\ &= \mathrm{P}(\mathscr{H}^{\{1\}}, t_1) \times \mathrm{P}(\mathscr{H}^{\{2\}}, t_2) \times \cdots \mathrm{P}(\mathscr{H}^{\{s\}}, t_s), \end{aligned} \tag{6.9}$$

where s is the number of rounds of repeated simulation realizations of epidemic invasion processes. Take the logarithm of Eq. (6.9), the joint likelihood function yields $\mathrm{L(P)} = \ln(\mathrm{P}(\mathscr{H}^{\{1\}}, t_1; \mathscr{H}^{\{2\}}, t_2; \cdots; \mathscr{H}^{\{s\}}, t_s))$.

Therefore, by means of the maximum likelihood estimation, we have $\frac{d\mathrm{L(P)}}{dp_{12}} = \frac{1}{p_{12}-1}(\sum_{i=1}^{s}(I_1^{\{i\}}(\eta_i) - \mathscr{H}^{\{i\}}) + \sum_{i=1}^{s}\sum_{j=1}^{\eta_i-1}I_1^{\{i\}}(j)) + \frac{1}{p_{12}}\sum_{i=1}^{s}\mathscr{H}^{\{i\}}$. Letting $\frac{d\mathrm{L(P)}}{dp_{12}} = 0$, we finally have

$$\hat{p}_{12} = \frac{\sum_{i=1}^{s}\mathscr{H}^{\{i\}}}{\sum_{i=1}^{s}\left[I_1^{\{i\}}(\eta_i) - \mathscr{H}^{\{i\}} + \mathscr{H}^{\{i\}} + \sum_{j=1}^{\eta_i-1}I_1^{\{i\}}(j)\right]} = \frac{\sum_{i=1}^{s}\mathscr{H}^{\{i\}}}{\sum_{i=1}^{s}\sum_{j=1}^{\eta_i}I_1^{\{i\}}(j)}, \tag{6.10}$$

where $\hat{p}_{12}$ represents the estimation of diffusion rate p_{12}.

6.3.2 The Case of a Meta-population Network

Mathematically, the estimation of diffusion rates requires the availability of a large number of epidemic realizations for a given meta-population network. However, the availability of such repeated data for emergent infectious diseases is rather limited in reality. Therefore, the estimation of diffusion rates in the general case of a meta-population network is infeasible due to the computational complexity and the limit of available data, which generally can be alternatively obtained from the statistics of public transportation section. The estimation of infection rate β in the general case of a meta-population network is addressed here.

Summing the number of infected individuals in Eq. (6.3) over all subpopulations i, we have $\sum_i \langle I_i(t+1) - I_i(t)\rangle = \sum_{i=1}^{N} \beta I_i(t) S_i(t)/N_i(t)$. Since $I_i(t) \ll N_i(t)$ at the early epidemic stage, it is simplified as $\sum_i \langle I_i(t+1) - I_i(t)\rangle \approx \beta \sum_i I_i(t)$. The term $I_i(t+1) - I_i(t)$ fluctuates around its mathematical expectation, and we have the approximation as

$$\beta \sim \frac{\sum_i (I_i(t+1) - I_i(t))}{\sum_i I_i(t)}. \tag{6.11}$$

Thus, given all recorded times $t_1, t_2, \ldots, t_{m'}$, the infection rate $\hat{\beta}$ is estimated as

$$\hat{\beta} = \arg\min_{\beta^*} \sum_{t=t_1}^{t_{m'}} |(I(t+1) - I(t)) - \beta^* I(t)|^2 = (X^\top X)^{-1} X^\top Y, \tag{6.12}$$

where $X^\top$ represents the transposition of X, and $X = [I(t_1), I(t_2), \ldots, I(t_{m'})]^{-1}$, $Y = [(I(t_1+1) - I(t_1)), (I(t_2+1) - I(t_2)), \ldots, (I(t_{m'}+1) - I(t_{m'}))]^{-1}$.

6.3.3 Example: Identifying the Diffusion Rate p_{12}

In this subsection, we only illustrate the identification performance of estimating diffusion rate p_{12} on a two-subpopulation SI model as an example. A more general case (in the sense of an arbitrary number of subpopulations) of example of identification performance of infection rate β will be investigated in Sect. 6.5. In the two-subpopulation case, statistic information of p_{12} is embedded in the surveillance infection data of the two subpopulations during the epidemic invasion process. As shown in Fig. 6.4, the estimation of p_{12} approaches the real value if the number of realizations increase, and the estimation error $|\hat{p}_{12} - p_{12}|$ is less than 5% of p_{12}. Finally the estimation of p_{12} as $\hat{p}_{12}$ tends to the real value.

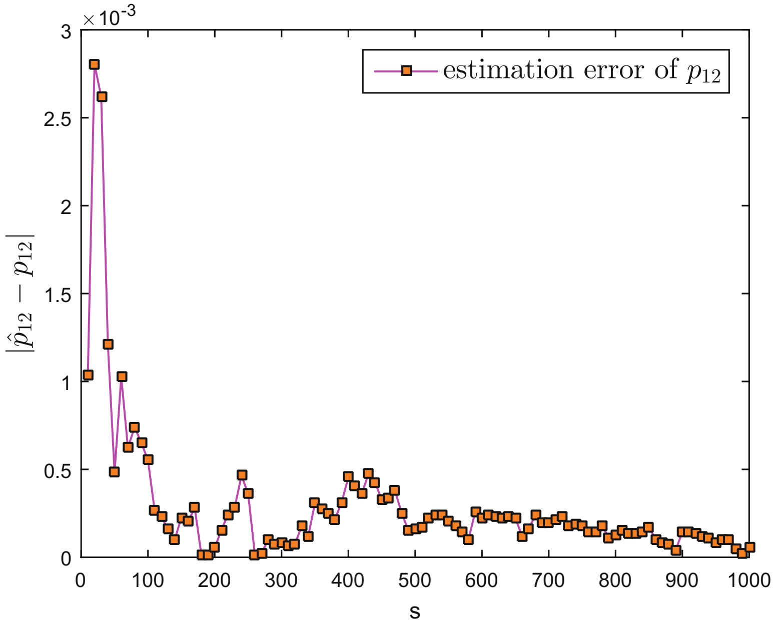

Fig. 6.4 The estimation error of diffusion rate p_{12} versus the number of realizations of the invasion process, and the error finally converges to zero. $\hat{p}_{12}$ is the estimated value of p_{12}. The actual value of diffusion rate p_{12} is 0.01

6.4 Identification of Invasion Pathways

During a real spatial cascade of an infectious disease, the spatial invasion pathways are the collection of directed transmission paths of an infectious disease rooted in the infected source subpopulation invading their susceptible neighbouring subpopulations. Actually, no one can predict such spatial invasion pathways to suppress the spreading processes at its infant prevalence. With the data availability of epidemic arrival time (EAT), i.e., the first invasion time discussed in the previous section, we may infer the patterns of invasion pathways.

Suppose one subpopulation is initially infected containing several infected individuals. As time evolves, the infected individuals of the seed subpopulation travel to the neighbouring subpopulations and try to infect their individuals. The successful invasion brings more invaded subpopulations with the cascade of infections. Therefore, the focus of interest is that when a subpopulation is invaded/infected by its $m(m \geq 2)$ infected neighbours with the available EAT data, how can we infer the culprit(s) and identify the invasion pathways in such a cascade infection? In the concerned situation, we assume that the surveillance infection data (the number of infected individuals of each subpopulation at each time t) is available as well as the topology of the meta-population network (including diffusion rates).

6.4.1 Invasion Partition and Types of Invasion Cases

We categorize all candidates of invasion pathways via the so-called invasion partition (INP) into four types of invasion cases (INCs), as shown in Fig. 6.5. An invasion case contains two sets, i.e., $\mathbb{S}$ and $\mathbb{I}$. Subpopulations which are not infected at $t_{EAT} - 1$ but infected at t_{EAT} are put in set $\mathbb{S}$, and their neighbours which are infected at $t_{EAT} - 1$ are put in set $\mathbb{I}$. All four types of invasion cases are defined.

(i) $I \mapsto S$: In this case, both $\mathbb{I}$ and $\mathbb{S}$ only have one subpopulation. That is to say, a susceptible subpopulation is infected at t_{EAT} by the first arrival of infected individual(s) from its unique neighbouring infected subpopulation at $t_{EAT} - 1$, and this infected subpopulation has no other newly infected neighbours at t_{EAT}.

(ii) $I \mapsto nS(n > 1)$: In this case, $\mathbb{I}$ contains one infected subpopulation, and $\mathbb{S}$ contains $n(n > 1)$ subpopulations. That is to say, an infected subpopulation

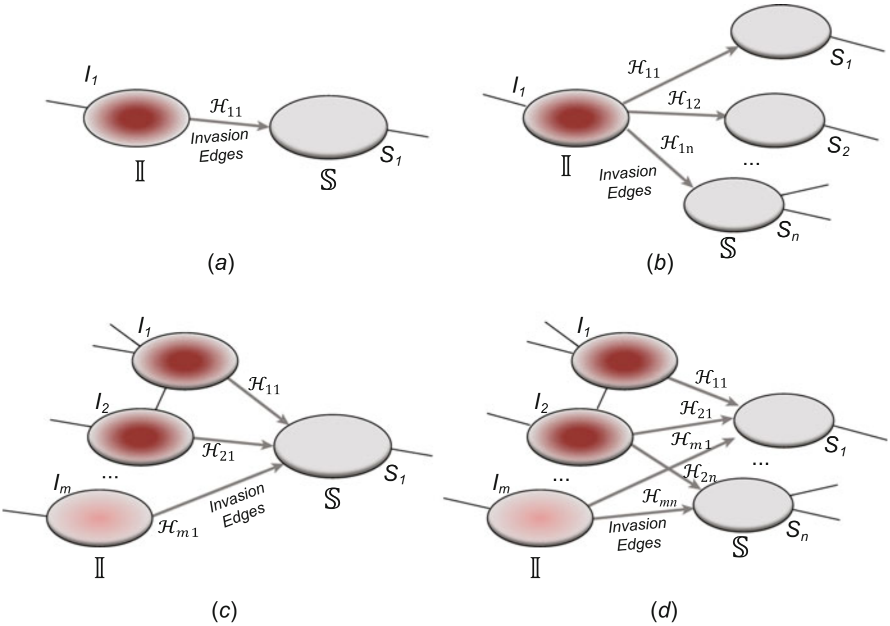

Fig. 6.5 (**a**) Example of $I \mapsto S$ INC, in which the infected individuals of only one infected subpopulation invades one susceptible subpopulation. The infected subpopulation is represented in *red*, while the plain patch is the subpopulation that remains susceptible before time t_{EAT} but will be infected between $t_{EAT} - 1$ to t_{EAT} due to the arrival of infected individuals from the upstream infected subpopulation. (**b**) Example of $I \mapsto nS$ INC, in which the infected individuals of only one infected subpopulation invades $n(n \geq 2)$ susceptible subpopulations. (**c**) Example of $mI \mapsto S$ INC, in which the infected individuals of m infected subpopulations invade one susceptible subpopulation. (**d**) Example of $mI \mapsto nS$ INC, in which the infected individuals of $m(m \geq 2)$ infected subpopulations invade $n(n \geq 2)$ susceptible subpopulations (From Wang et al. [35])

simultaneously infects its $n(n > 1)$ susceptible neighbours, each of which has only one infected neighbouring subpopulation.

(iii) $mI \mapsto S(m > 1)$: In this case, $\mathbb{I}$ consists of $m(m > 1)$ subpopulations, and $\mathbb{S}$ only contains one single subpopulation. That is to say, a susceptible subpopulation is infected by the first arrival of infected individual(s) coming from its $m(m > 1)$ infected neighbouring subpopulation, which has no other newly infected neighbours at this time.

(iv) $mI \mapsto nS(m, n > 1)$: In this case, sets $\mathbb{S}$ and $\mathbb{I}$ both contain more than one subpopulation. The edges from $\mathbb{I}$ to $\mathbb{S}$ form a connected subgraph. Each previously susceptible subpopulation in $\mathbb{S}$ is infected by the new arrival of infected individual(s) from at least one of the m infected subpopulations in $\mathbb{I}$. Each subpopulation in $\mathbb{I}$ has no other newly infected neighbours except the susbpopulations in $\mathbb{S}$ at this time.

Figure 6.5 illustrates such four types of invasion cases as $I \mapsto S$, $mI \mapsto nS(n > 1)$, $mI \mapsto S(m > 1)$ and $mI \mapsto nS(m, n > 1)$. Besides, we define the directed edges from infected subpopulation i in $\mathbb{I}$ to susceptible subpopulations in $\mathbb{S}$ as invasion edges, which are the candidates of invasion pathways. Therefore, we define a decomposition procedure *invasion partition* (INP) to achieve the task of dividing subpopulations and edges into such invasion cases. As summarized in Algorithm 1, we propose a heuristic algorithm to achieve the INP task.

Algorithm 1 Invasion Partition (INP)

```
1: At an epidemic arrival time, collect all newly infected subpopulations as initial $\mathbb{S}$ and their previously infected neighbours as $\mathbb{I}$;
2: Start with an arbitrary element $S_i$ in set $\mathbb{S}$, to compose the initial $\mathbb{S}^*$;
3: Find all neighbors of $S_i$ in set $\mathbb{I}$ to compose the set $\mathbb{I}^*$ ;
4: For each new member in $\mathbb{I}^*$, find its new neighbours in the $\mathbb{S}$ to update $\mathbb{S}^*$ if any;
5: For each new member in $\mathbb{S}^*$, find its new neighbours in the $\mathbb{I}$ to update $\mathbb{I}^*$ if any;
6: Repeat the above two steps until we cannot find any new neighbours in $\mathbb{S}$ and $\mathbb{I}$, we get an invasion case consisting of $\mathbb{I}^*$ and $\mathbb{S}^*$, then update the $\mathbb{S}$ and $\mathbb{I}$;
7: Repeat steps 2–6 to get new invasion cases until there are no elements in $\mathbb{S}$.
```

6.4.2 Observability of a Subpopulation and an Edge

We further classify the observability of a subpopulation and an edge. Observability of a subpopulation is defined by comparing the number of infected individuals of subpopulation i at time $t_{EAT} - 1$ and t_{EAT}, which reflects the information held for the inference of relevant invasion pathway. Observability of an directed edge emanated from an infected subpopulation can be defined by the types of subpopulations it connects to.

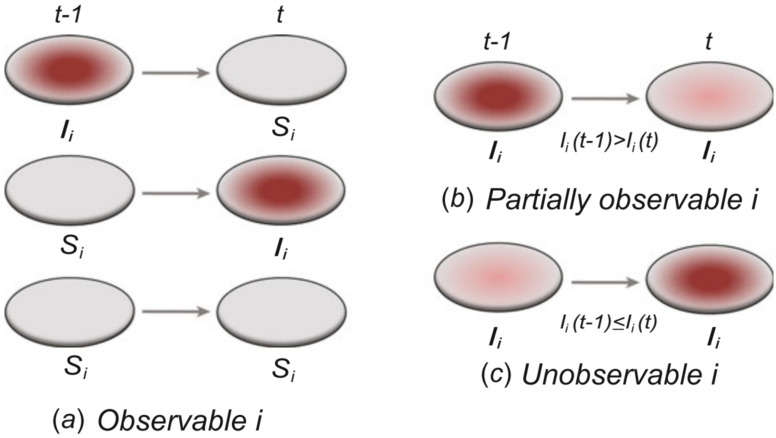

Fig. 6.6 Illustration of subpopulation observability: (**a**) observable subpopulations, (**b**) partially observable subpopulation, and (**c**) unobservable subpopulation. Here time t is t_{EAT} (From Wang et al. [35])

(i) *Observable Subpopulation:* From $t_{EAT} - 1$ to t_{EAT}, subpopulation i is an observable subpopulation if it experiences one of the following three state transitions. The first is $S_i \rightarrow I_i$, which indicates that this subpopulation has been infected (for the first time) during this period by infected individuals (because $I_i(t)$ is available). The second is $I_i \rightarrow S_i$. We know how many infected individuals diffused from this subpopulation in this case. The third is $S_i \rightarrow S_i$. This case represents subpopulation i keeps its susceptible status.

(ii) *Partially Observable Subpopulation:* The number of infected individuals of an infected subpopulation may decrease, that is to say $I_i(t_{EAT}) < I_i(t_{EAT} - 1)$ and $I_i(t_{EAT}) > 0$. We call subpopulation i is a partially observable subpopulation, because we know at least $\Delta I_i(t_{EAT}) = |I_i(t_{EAT}) - I_i(t_{EAT} - 1)|$ infected individuals leave i.

(iii) *Unobservable Subpopulation*: If the number of infected individuals does not decrease, i.e., $I_i(t_{EAT}) \geq I_i(t_{EAT} - 1)$, it is difficult to judge whether and how many infected hosts leave subpopulation i. We call it unobservable subpopulation.

Here the observability of a subpopulation indicates the diffusion information of this subpopulation. Figure 6.6a–c illustrate the above cases. Together with the observability of a subpopulation, the directed edges emanated from an infected subpopulation (here denoted i) in set $\mathbb{I}$ can be classified into three types, i.e., observable edges, partially observable edges and unobservable edges:

(i) *Observable Edges:* Any directed edge from i to observable subpopulation j whose transition is $S_j \rightarrow S_j$ or $I_j \rightarrow S_j$ from $t_{EAT} - 1$ to t_{EAT}. This edge implies no infected hosts move from i.

(ii) *Partially Observable Edges:* If an directed edge emanated from infected subpopulation i to a partially observable subpopulation, the edge is partially observable.

(iii) *Unobservable Edges:* If infected subpopulation i connects with an unobservable subpopulation, this directed edge from i is an unobservable edge.

6.4.3 Accurate Identification of Invasion Pathways

We now consider to accurately identify the invasion pathways. Among the four types of invasion cases (INCs), since the two types of INCs ($I \mapsto S$ and $I \mapsto nS$, $n \geq 2$) have the unique invasion edge(s) from the neighboring infected subpopulation, the invasion pathways therefore are easy to identify accurately. We only need concern the other two types of INCs, i.e., $mI \mapsto S$ and $mI \mapsto nS$.

6.4.3.1 The Case of $mI \mapsto S$ ($m > 1$)

A representative $mI \mapsto S(m > 1)$ INC (Fig. 6.5c) consists of two sets. Set $\mathbb{I} = \{I_1, I_2, \ldots, I_m\}$ is composed of the infected subpopulations at $t_{EAT} - 1$, and set $\mathbb{S} = \{S_1\}$ is composed of the susceptible subpopulation(s) at $t_{EAT} - 1$ which are infected at t_{EAT}. Assume subpopulation S_1 is infected at t_{EAT} by the first arrival of $\mathscr{H}$ infected individuals coming from some of the infected subpopulations in $\mathbb{I}$, where $\mathscr{H}$ is a positive integer.

Suppose $\mathscr{H}_{i1}$ is the actual number of infected individuals travelling from an infected subpopulation I_i in set $\mathbb{I}$, and we have

$$\sum_{i=1}^{m} \mathscr{H}_{i1} = \mathscr{H}, \tag{6.13}$$

where $0 \leq \mathscr{H}_{i1} \leq \mathscr{H}$, $\mathscr{H}_{i1} \leq I_i(t_{EAT} - 1)$, and $0 \leq i \leq m$. $\mathscr{H}$ is available from the infection data, while we do not know $\mathscr{H}_{i1}$. To reach the unique solution of Eq. (6.13) which corresponds to a set of invasion pathways of $mI \mapsto S(m > 1)$, we give Theorem 1 to accurately identify the invasion pathways of INC $mI \mapsto S(m > 1)$.

Theorem 1 *The invasion pathways of the invasion case $mI \mapsto S(m > 1)$ can be accurately identified, given the following two conditions are satisfied: (1) among m possible sources illustrated in set $\mathbb{I}$, there are only $m'(m' \leq m)$ partially observable subpopulations $\mathbb{I}'$, whose neighbouring subpopulations j (excluding the invasion destination S_1) only experience the transition $S_j \rightarrow S_j$ or $I_j \rightarrow S_j$ at that EAT, (2) $\sum_{i \in \mathbb{I}'} \left[I_i(t_{EAT} - 1) - I_i(t_{EAT}) \right] = \mathscr{H}$.*

Proof According to the definition of observability, in an INC, the number of local infected individuals in an partially observable source i will be decreased by

$\big[I_i(t_{EAT} - 1) - I_i(t_{EAT})\big]$ due to the movement of infected individuals. If the subpopulations j in the neighbourhood of i only experience the transition of $S_j \rightarrow S_j$ or $I_j \rightarrow S_j$ from $t_{EAT} - 1$ to t_{EAT}, they do not to receive the infected individuals from subpopulation i. Therefore, the newly infected subpopulation S_1 is the only destination for those infected individuals departing from the partially observable sources. Since $m' \leq m$, the second condition guarantees that Eq. (6.13) only has a unique solution, which corresponds to the accurate identification of invasion pathways of this invasion case. □

6.4.3.2 The Case of $mI \mapsto nS(m > 1, n > 1)$

The final typical INC $mI \mapsto nS$ as shown in Fig. 6.5d includes set $\mathbb{I} = \{I_i|i = 1, 2, \ldots, m\}$ and $\mathbb{S} = \{S_i|i = 1, 2, \ldots, n\}$. Denote $\{\mathscr{H}_i|i = 1, 2, \ldots, n\}$ the number of the first arrival of infected individuals to susceptible subpopulation S_i in set $\mathbb{S}$, and $U_i(i = 1, 2, \ldots, m)$ the subset of susceptible neighbouring subpopulations in set $\mathbb{S}$ of infected subpopulation I_i , and $Y_j(j = 1, 2, \ldots, n)$ the subset of infected neighbouring subpopulations in set $\mathbb{I}$ of susceptible subpopulation S_j.

Define $\sigma = \{\{\mathscr{H}_{i1}|i \in Y_1\}, \ldots, \{\mathscr{H}_{in}|i \in Y_n\}\}$ as a potential solution for the $mI \mapsto nS$, if σ is subject to the following two conditions: (i)

$$\sum_{i \in Y_k} \mathscr{H}_{ik} = \mathscr{H}_k, \tag{6.14}$$

where $\mathscr{H}_{ik}(\geq 0)$ is the number of infected hosts invading subpopulation S_k from I_i at t_{EAT}; (ii) For any $\mathscr{H}_{ik}$, we have $\sum_{k \in U_i} \mathscr{H}_{ik} \leq I_i(t_{EAT} - 1)$, where $1 \leq i \leq m, 1 \leq k \leq n$.

Suppose an $mI \mapsto nS$ has M potential solutions, and $\sigma_j = \{\{\mathscr{H}_{i1}^{(j)}|i \in Y_1\}, \ldots, \{\mathscr{H}_{in}^{(j)}|i \in Y_n\}\}$ $(1 \leq j \leq M)$ represents one of the solutions.

Given some specific prerequisites (as the conditions of Theorem 2), Eq. (6.14) has a unique solution, which implies that the invasion pathway(s) can be identified accurately. Theorem 2 elucidates this scenario.

Theorem 2 *The invasion pathway(s) of the invasion case $mI \mapsto nS(m, n > 1)$ can be identified accurately, given the following three conditions are satisfied: (1) the number of invasion edges $E_{in} \leq n + m$, (2) the neighbouring subpopulations j of each subpopulation in set $\mathbb{I}$ are with the transition $S_j \rightarrow S_j$ or $I_j \rightarrow S_j$ except their neighbouring subpopulations in set $\mathbb{S}$ during $t_{EAT} - 1$ to t_{EAT}, (3) $\sum_{i=1}^{m} \Delta I_i(t_{EAT}) = \sum_{k=1}^{n} \mathscr{H}_k$.*

Proof Since the number of infected individuals in the partially observable subpopulation i reduces at time t_{EAT}, i.e., $I_i(t_{EAT}) < I_i(t_{EAT} - 1)$, $I_i(t_{EAT}) > 0$, it is inevitable that a few infected individuals diffuse away from subpopulation i. Occurring the state transitions of $S_j \rightarrow S_j$ or $I_j \rightarrow S_j$ from $t_{EAT} - 1$ to t_{EAT}, subpopulations j in the neighbourhood of i (excluding the new infected subpopulation j) cannot receive infected individuals. Therefore, the only possible destination for those infected individuals is subpopulation S_k in $\mathbb{S}$.

The conditions $E_{in} \leq n+m$ and $\sum_{i=1}^{m} \Delta I_i(t_{EAT}) = \sum_{k=1}^{n} \mathscr{H}_k$ make the equations $\sum_{i\in Y_k} \mathscr{H}_{ik} = \mathscr{H}_k$ and $\sum_{k\in U_i} \mathscr{H}_{ik} = \Delta I_i(t_{EAT})$ have the unique solution $\sigma = \{\{\mathscr{H}_{i1}|i \in Y_1\}, \ldots, \{\mathscr{H}_{in}|i \in Y_n\}\}$. The reason is that rank(A_{coef})=E_{in}, where A_{coef} is the coefficient matrix of equations $\sum_{i\in Y_k} \mathscr{H}_{ik} = \mathscr{H}_k$ and $\sum_{k\in U_i} \mathscr{H}_{ik} = \Delta I_i(t_{EAT})$. Thus the invasion pathway(s) of this $mI \mapsto nS(m, n > 1)$ can be identified accurately. □

6.4.4 Identification for Potential Invasion Pathways

Now we are in the position to construct the whole framework of identifying invasion pathways, namely, the invasion pathways identification (IPI) algorithm as summarized as below.

(i) Invasion partition: $\mathrm{T}_{\text{whole invasion pathways}}$ is defined as the whole invasion pathways of an invasion process. At each EAT, we get four types of invasion cases (i.e. $I \mapsto S, I \mapsto nS, mI \mapsto S, mI \mapsto nS(m > 1, n > 1)$). Suppose $\mathrm{T}_{\text{whole invasion pathways}}$ is contained in all Λ INCs. Denote by $\hat{a}_i$ the identified invasion pathways of INC_i, which can be optimally solved by (stochastic) dynamic programming as

$$\mathrm{T}_{\text{whole invasion pathways}} = \text{opt} \sum_{i=1}^{\Lambda} \hat{a}_i, \tag{6.15}$$

where "opt" represents the optimal solution via dynamic programming.

(ii) Accurate identification: For the two cases of $I \mapsto S$, $I \mapsto nS$, it is easy to reach the accurate identification of invasion pathways. In the other two cases of $mI \mapsto S, mI \mapsto nS$, we first evaluate whether $mI \mapsto S$ or $mI \mapsto nS$ can be accurately identified or not. If yes, Theorems 1 and 2 work out the accurate identification.

(iii) Identification of potential invasion pathways: If accurate identification is not feasible, we propose an efficient optimization method based on the maximum likelihood estimation to identify the most likely invasion pathways. We define the maximum likelihood (ML) estimator as

$$\hat{a}_i = \underset{a_i \in INC_i}{\arg\max}\ P(a_i|INC_i), \tag{6.16}$$

where $P(a_i|INC_i)$ is the likelihood of uncovering the potential pathway a_i, supposing the actual pathway is a_i^*. Therefore, we evaluate $P(a_i|INC_i)$ and choose the maximal likelihood one as a_i^* from all potential pathways $a_i \in INC_i$.

(iv) The whole spatial invasion pathways can be reconstructed by assembling all invasion cases chronologically.

Therefore, in the situations where accurate identification of invasion pathways is not feasible, e.g., the conditions of Theorems 1 and 2 are not satisfied, Eqs. (6.13) and (6.14) may have a number of potential solutions which correspond to a set of potential invasion pathways. Therefore, we propose the identification algorithm to infer the most likely pathways among all potential invasion pathways. Herein we unify $mI \mapsto S(m > 1)$ and $mI \mapsto nS(m > 1, n > 1)$ as $mI \mapsto nS(m > 1, n \geq 1)$.

Denote $\Omega(\mathscr{H}^{(j)}_{kk_\hbar})$ the transfer estimator of infected subpopulation I_k in $\mathbb{I}$, $k_\hbar \in Y_k$. Here the transfer estimator is used to estimate the diffusion likelihood if I_k diffuses $\mathscr{H}^{(j)}_{kk_\hbar}$ infected individuals to $S_{k_\hbar}$. Thus, the likelihood of potential solution σ_j of an INC $mI \mapsto nS(m > 1, n \geq 1)$ is presented by

$$P(\sigma_j|INC_{mInS}) = \prod_{k=1}^{m} \Omega(\mathscr{H}^{(j)}_{kk_\hbar}) \Big/ \sum_{i=1}^{M} \prod_{k=1}^{m} \Omega(\mathscr{H}^{(i)}_{kk_\hbar}), \tag{6.17}$$

where M represents the number of solution σ_j.

We now consider the events from $t_{EAT} - 1$ to t_{EAT}, and give some definitions. We assume an infected subpopulation I_i in $\mathbb{I}$ emanates k_i edges in total, among which there are $\rho_i (1 \leq \rho_i \leq n)$ invasion edge(s) labeled as $1, 2, \ldots, \rho_i$ with the corresponding diffusion rates $p_\hbar, \hbar \in [1, \rho_i]$, $\hbar$ is an integer. We suppose $\mathscr{H}_{ii_\hbar}$ infected hosts invade its neighbouring subpopulations in the subset $\{Y_i = i_\hbar\}$ at t_{EAT}. Assume there are ℓ_i unobservable and partially observable edges, labelled as $1 + \rho_i, \ldots, \ell_i + \rho_i$. Along each unobservable or partially observable edge, the traveling rate is p_ℓ, $\ell \in [1, \ell_i]$, and x_ℓ infected hosts leave I_i. Accordingly, in total $\eta_i = \sum_\ell x_\ell$ infected individuals leave I_i through the unobservable and partially observable edges. Now there remain $k_i - \ell_i - \rho_i$ observable edges, labelled as $\ell_i + \rho_i + 1, \ldots, k_i$. Along each observable edge, the diffusion rate is $p_\aleph$, integer $\aleph \in [\ell_i + \rho_i + 1, k_i]$, and $x_\aleph$ infected individuals leave I_i. With probability $\overline{p_i} = 1 - \sum_\hbar p_\hbar - \sum_\ell p_\ell - \sum_\aleph p_\aleph$, an infected individual keeps staying at subpopulation I_i. There are $\overline{x_i}$ infected individuals staying in subpopulation I_i with probability $\overline{p_i}$. Because I_i connects the unobservable and partially observable infected subpopulations, we obtain $\sum_\ell x_\ell + \overline{x_i} = \eta'$.

Therefore, we have the transfer likelihood estimator Ω of I_i in the following three parts.

(a) *Unobservable Subpopulation I_i:* It is difficult to estimate whether and how many infected hosts move to which neighbours due to $\Delta I_i(t_{EAT}) = I_i(t_{EAT} - 1) - I_i(t_{EAT}) \leq 0$ (we have $I_i(t_{EAT} - 1) \leq I_i(t_{EAT})$ because unobservable subpopulation I_i). We write the transfer likelihood estimator of I_i as

$$\begin{aligned}\Omega_u(\mathscr{H}_{ii_\hbar}) = P(&\mathscr{H}_{ii_\hbar}, p_\hbar, \hbar = [1, 2, \ldots, \rho]; x_\ell, p_\ell, \ell = [1 + \rho, 2 + \rho, \\ &\ldots, l + \rho]; x_\aleph, p_\aleph, \aleph = [l + \rho + 1, l + \rho + 2, \ldots, k]; \overline{x_i}, \overline{p_i}).\end{aligned} \tag{6.18}$$

With the definition of observable edges, the transfer likelihood estimator is simplified as

$$\Omega_u = \frac{I_i(t-1)!}{\prod_{\hbar} \mathscr{H}_{ii\hbar}! \eta_i'!} \prod_{\hbar} p_{\hbar}^{H_{ii\hbar}} \Big[\sum_{\ell} p_\ell + \overline{p_i} \Big]^{\eta_i'}. \tag{6.19}$$

(b) Observable Subpopulation I_i ($I_i \rightarrow S_i$): Given an $I \rightarrow S$ observable subpopulation I_i, the infected individuals $H_i = \{\mathscr{H}_{ii\hbar} | \hbar = 1, 2, \ldots, \rho\}$ moved out of subpopulation I_i to $S_{i\hbar}$ are all from the term of $\Delta I_i(t_{EAT})$. Therefore, its transfer likelihood estimator is derived as

$$\Omega_{ob} = \frac{\Delta I_i(t)!}{\prod_{\hbar} \mathscr{H}_{ii\hbar}!(\Delta I_i(t) - \sum_{\hbar} \mathscr{H}_{ii\hbar})!} \prod_{\hbar} (\frac{p_{\hbar}}{\sum_{k=1}^{l+\rho} p_k})^{\mathscr{H}''_{ii\hbar}} \cdot$$
$$(\frac{\sum_{\ell} p_\ell}{\sum_{j=1}^{l+\rho} p_j})^{\Delta I_i(t) - \sum_{\hbar} \mathscr{H}_{ii\hbar}}, \tag{6.20}$$

where $\Delta I_i(t_{EAT}) = I_i(t_{EAT} - 1) - I_i(t_{EAT}) = I_i(t_{EAT} - 1)$ (we have $I_i(t_{EAT}) = 0$ because observable subpopulation I_i ($I_i \rightarrow S_i$)).

(c) Partially Observable Subpopulation I_i: Because $\Delta I_i(t_{EAT}) = I_i(t_{EAT} - 1) - I_i(t_{EAT}) > 0$, at least $\Delta I_i(t_{EAT})$ infected individuals leave subpopulation I_i from $t_{EAT} - 1$ to t_{EAT} according to the definition of partially observable subpopulation. $H_i = \{\mathscr{H}_{ii\hbar} | \hbar = 1, \ldots, \rho\}$ is decomposed into two subsets: $H_i' = \{\mathscr{H}'_{ii\hbar} | \hbar = 1, \ldots, \rho\}$ and $H_i'' = \{\mathscr{H}''_{ii\hbar} | \hbar = 1, 2, \ldots, \rho\}$, $\mathscr{H}'_{ii\hbar} + \mathscr{H}''_{ii\hbar} = \mathscr{H}_{ii\hbar}$, where $\mathscr{H}'_{ii\hbar} \geq 0, \mathscr{H}''_{ii\hbar} \geq 0$. $H_i' = \{\mathscr{H}'_{ii\hbar} | \hbar = 1, \ldots, \rho\}$ represents the set of infected individuals departing from $I_i(t_{EAT} - \Delta t) - \Delta I_i(t_{EAT})$, and $H_i'' = \{\mathscr{H}''_{ii\hbar} | \hbar = 1, \ldots, \rho\}$ denote the infected individuals departing from $\Delta I_i(t_{EAT})$. We then have the transfer likelihood estimator in the following two cases.

Case 1: $\sum_{\hbar} \mathscr{H}_{ii\hbar} \geq \Delta I_i(t_{EAT})$

The transfer likelihood estimator is

$$\Omega_{pu} = \sum_{\phi=0}^{\Delta I_i(t_{EAT})} \sum_{\sum \mathscr{H}''_{ii\hbar} = \phi} P_1 P_2, \tag{6.21}$$

where

$$P_1 = \frac{\Delta I_i(t_{EAT})!}{\prod_{\hbar} \mathscr{H}''_{ii\hbar}!(\Delta I_i(t_{EAT}) - \phi)!} \prod_{\hbar} (\frac{p_{\hbar}}{\sum_{k=1}^{l+\rho} p_k})^{\mathscr{H}''_{ii\hbar}} (\frac{\sum_{\ell} p_\ell}{\sum_{j=1}^{l+\rho} p_j})^{\Delta I_i(t_{EAT}) - \phi},$$

$$P_2 = \frac{(I_i(t_{EAT} - \Delta t) - \Delta I_i(t_{EAT}))!}{\prod_{\hbar} \mathscr{H}'_{ii\hbar}!(I_i(t_{EAT} - \Delta t) - \Delta I_i(t_{EAT}) - \sum_{\hbar} \mathscr{H}_{ii\hbar} + \phi)!} \prod_{\hbar} p_{\hbar}^{\mathscr{H}'_{ii\hbar}}$$
$$\times (\sum_{\ell} p_\ell + \overline{p_i})^{I_i(t_{EAT} - \Delta t) - \Delta I_i(t_{EAT}) - \sum_{\hbar} \mathscr{H}_{ii\hbar} + \phi}.$$

Here, $\phi = \sum_{\hbar} \mathscr{H}''{}_{ii_{\hbar}} (0 \leq \phi \leq \Delta I_i(t_{EAT}))$, which represents the sum of infected individuals travelling from subpopulation I_i to $S_{i_{\hbar}}$. For a given ϕ, we need to enumerate all possible sets $H_i'' = \{\mathscr{H}_{ii_j}''|j = 1, \ldots, \rho\}$ to calculate the Ω_{pu}.

Case 2: $\sum_{\hbar} \mathscr{H}_{ii_{\hbar}} < \Delta I_i(t)$

Denote $\phi = \sum_{\hbar} \mathscr{H}''{}_{ii_{\hbar}} (0 \leq \phi \leq \sum_{\hbar} \mathscr{H}_{ii_{\hbar}})$. Similar to Case 1, we should enumerate all possible permutations of $H_i'' = \{\mathscr{H}_{ii_j}''|j = 1, \ldots, \rho\}$ for a fixed ϕ. Therefore, in this case we have the transfer likelihood estimator of I_i as

$$\Omega_{pu} = \sum_{\phi=0}^{\sum_{\hbar} \mathscr{H}_{ii_{\hbar}}} \sum_{\sum \mathscr{H}_{ii_{\hbar}}'' = \phi} P_1 P_2, \tag{6.22}$$

where P_1 and P_2 are the same as those in Eq. (6.21).

According to Eq. (6.17), the most likely invasion pathways for an INC $mI \mapsto nS (m > 1, n \geq 1)$ are identified as

$$\hat{a}^{mI \mapsto nS} = \arg\max_{\sigma_i} P(\sigma_i | INC_{mI \mapsto nS}) = \arg\max_{a_i} P(a_i | INC_{mI \mapsto nS}). \tag{6.23}$$

If the number of the first arrival infected individuals $\mathscr{H}_{ij} \geq 3$, multiple potential solutions may correspond to the same set of potential pathway(s). In this case, we merge the transfer likelihood of all potential solutions of this INC if they belong to the same invasion pathways. Then we find out the most likely invasion pathways corresponding to the maximum likelihood.

After identifying the potential invasion pathways, the whole invasion pathway $\mathrm{T}_{\text{whole invasion pathways}}$ can be reconstructed chronologically by assembling all INCs. Finally, we depict the IPI algorithm explicitly with the pseudocodes as outlined in Algorithm 2.

Algorithm 2 Invasion Pathways Identification (IPI)

```
Inputs: the time series of infection data $I_i(t)$ and topology of network $G$
Find all EAT data
for each EAT
  Invasion partition to find out the $I \mapsto S$, $I \mapsto nS$, $mI \mapsto S$ and $mI \mapsto nS$.
  for each $mI \mapsto S$ or $mI \mapsto nS$
    if it satisfies conditions of Th 1 or Th 2
      Compute the unique invasion pathway
    else It does not satisfy conditions of Th 1 or Th 2
      Find all $M$ potential solutions $\sigma_j$
      Compute the $P(\sigma_j|INC_{mI \mapsto S})$ or $P(\sigma_j|INC_{mI \mapsto nS})$
      Merge the $P(\sigma_j|INC_{mI \mapsto S})$ or $P(\sigma_j|INC_{mI \mapsto nS})$ of $\sigma_j$ corresponding to same pathway(s)
    end if
  end for
  Find invasion pathway $a^{mI \mapsto S}$ or $a^{mI \mapsto nS}$ that maximize $P(\sigma_j|INC_{mI \mapsto S})$ or $P(\sigma_j|INC_{mI \mapsto nS})$
end for
Reconstruct the whole invasion pathways (T) by assembling each invasion case chronologically
```

6.4.5 Identifiability of Invasion Pathways

We now evaluate the identifiability of invasion pathways of all invasion cases. Denote π the likelihood corresponding to the most likely pathways for a given invasion case. Therefore we have

$$\pi(\sigma) = \sup_{\sigma_i}\{P(\sigma_i|INC)\}. \tag{6.24}$$

Property 1 Given an invasion case '$mI \mapsto S$' or '$mI \mapsto nS$', $P(\sigma_j|INC) = \frac{\prod_{k=1}^{m}\Omega}{\sum_{i=1}^{M}\prod_{k=1}^{m}\Omega}$, there must exist P_{min} and P_{max} satisfying

$$P_{min} \leq \pi(\sigma) \leq P_{max}. \tag{6.25}$$

Proof Suppose that $P(\sigma_1|INC) \leq \ldots \leq P(\sigma_M|INC)$, where M is the number of potential solutions. Thus $P_{max} = (P(\sigma_M|INC)/P(\sigma_2|INC) + \ldots + P(\sigma_M|INC))$; Because $\pi(\sigma) \geq 1/M$, let $P_{min} = max\{1/M, P(\sigma_M|INC)/(P(\sigma_1|INC) + \sum_{j=1}^{M} P(\sigma_j|INC))\}$. We have $P_{min} \leq \pi(\sigma) \leq P_{max}$. □

We define an entropy to characterize the likelihood vector of M potential pathways of an INC.

Definition 1 (Entropy of Likelihoods of M Potential Solutions) Define the normalized entropy of transfer likelihood $P(\sigma_1|INC), \ldots, P(\sigma_M|INC)$ as

$$\mathscr{S} = -\frac{1}{\log M}\sum_{i=1}^{M} P(\sigma_i|INC)\log P(\sigma_i|INC). \tag{6.26}$$

This likelihood entropy $\mathscr{S}$ tells the information embedded in the likelihood vector of the potential solutions of a given INC.

Definition 2 (Identifiability of Invasion Pathways) Define the identifiability of invasion pathways to characterize the feasibility to identify an invasion case as

$$\Pi = \pi(\sigma)(1 - \mathscr{S}). \tag{6.27}$$

Definition 2 tells that the bigger $\pi(\sigma)$ and the smaller entropy $\mathscr{S}$, the easier to identify the epidemic invasion pathways for an invasion case.

6.4.6 Examples

We illustrate the performance of our proposed IPI algorithm to identify the invasion pathways, with the maximal connected component of the American airports network (AAN, Fig. 6.7) to form a meta-population network. Note that the data to construct

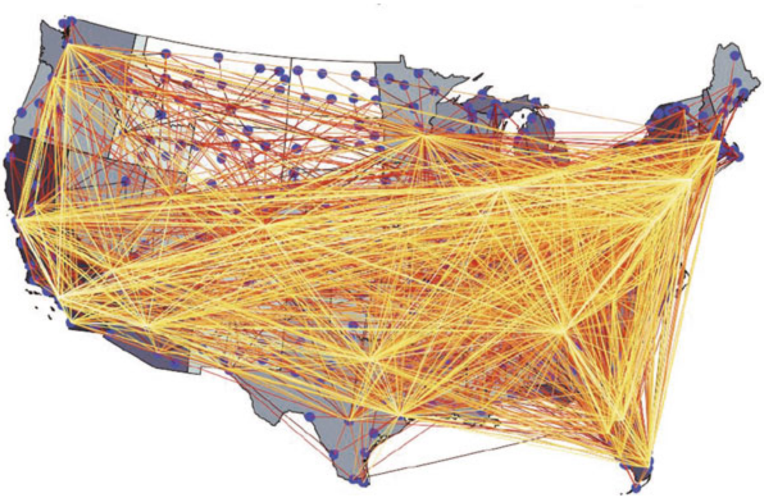

Fig. 6.7 Illustration of an American airports network (From Brockmann et al. [42])

the AAN was collected from the U.S. demographic statistical data and domestic air transportation [35, 43]. Here, the AAN is a weighted and directed graph having $V = 404$ nodes (airports) and $E = 6480$ weighted and directed edges representing flight routes. The weight of edge E_{ij} is defined as diffusion rate $p_{ij} = \frac{\langle w_{ij} \rangle}{\langle N_i \rangle}$, where $\langle w_{ij} \rangle$ is the daily amount of passengers of the flight from i to j, $\langle N_i \rangle$ is the population of serving areas [43] of airport i. The average degree of the AAN is $\langle k \rangle \approx 16$, and the range of degree k is [1,158]. The range of distributions of $\langle w_{ij} \rangle$ and p_{ij} is $[1, 9100]$ and $[7.4 \times 10^{-8}, 0.03]$, respectively. The range of distribution of $\langle N_i \rangle$ is $[6100, 1.907 \times 10^7]$, and the total population of the AAN is $N_{total} \approx 0.243 \times 10^9$, i.e., approximately the whole population of the United States of America. Therefore, the AAN as the sample of a meta-population network shows high heterogeneity of connectivity patterns, traffic capacities as well as the population distribution [43].

To verify the performance of the proposed IPI algorithm, we select three methods [15, 31, 32] as the benchmark for comparison, which generate the shortest path trees or minimum spanning trees of a meta-population network. In more detail, [31] generates the average-arrival-time-based (ARR) shortest path tree, and [15] generates the effective-distance-based (EFF) most probable paths, and [32] generates the Monte-Carlo-Maximum-Likelihood-based (MCML) most likely epidemic invasion tree.

We define the identifying accuracy as the ratio of the number of correctly identified invasion pathways by each method to the number of true invasion pathways. We also compute the accuracy of accumulative INCs of $mI \mapsto S$ and $mI \mapsto nS$, which is defined as the ratio of the number of correctly identified invasion pathways by each method to the number of true invasion pathways in this INC. Besides, we also make the comparison of the identification accuracy at the

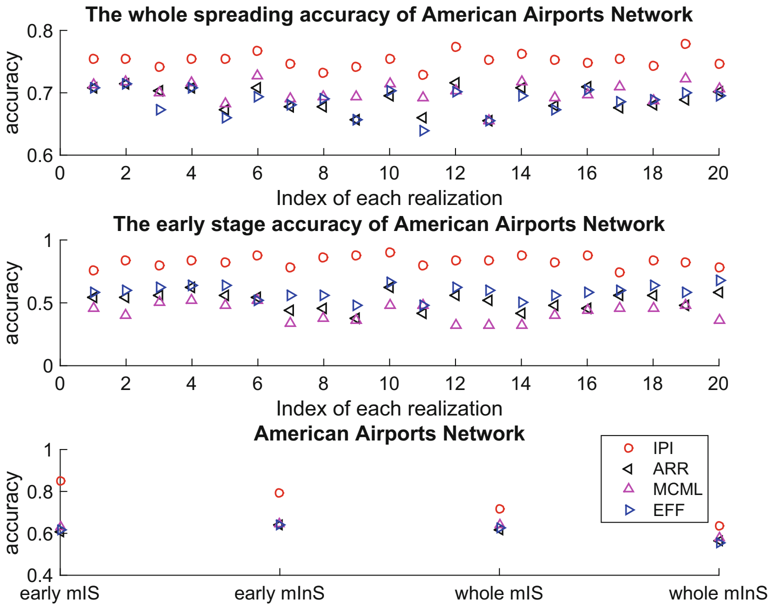

Fig. 6.8 (*Top*) The wholly identifying accuracy of the invasion pathways on the AAN with 20 rounds of independent realizations. (*Middle*) The identifying accuracy of the invasion pathways for the early stage (before infecting 50 subpopulations) on the AAN with 20 rounds of independent realizations. (*Bottom*) The accumulative identifying accuracy of invasion cases ($mI \mapsto S$ and $mI \mapsto nS$) for the early stage and the whole invasion pathways on the AAN. Here "mIS" and "mInS" stand for $mI \mapsto S$ and $mI \mapsto nS$, respectively (From Wang et al. [35])

early stage of epidemic dynamics, which is defined as the period when the first 50 subpopulation have been infected. In the top and middle panels of Fig. 6.8, we observe the whole identification accuracy and the early-stage identification accuracy, while the bottom panel of Fig. 6.8 presents the early and whole accumulative identification accuracy of $mI \mapsto S$ and $mI \mapsto nS$ through 20 independent realizations on the AAN, respectively. Here the whole identification accuracy means the identification accuracy of whole meta-population network has been infected. The seed subpopulation in all such independent realizations is set as the Sun Valley Airport in Bullhead City, Arizona. We clearly observe that the IPI algorithm is more accurate at identifying the invasion pathways than other benchmark methods.

We then visualize the identified invasion pathways (the lower panel of Fig. 6.9) during the early stage of a realization compared with the actual invasion pathways (the upper panel of Fig. 6.9). The weights (diffusion rates) of invasion edges are shown by the thicknesses of lines, and arrows represent the directions of invasions. We observe that most of the invasion pathways are correctly identified to form the invasive backbone of this realization of an epidemic dynamics, while there still exist some wrongly identified pathways in some INCs, indicating the necessity of defining the identifiability of an INC.

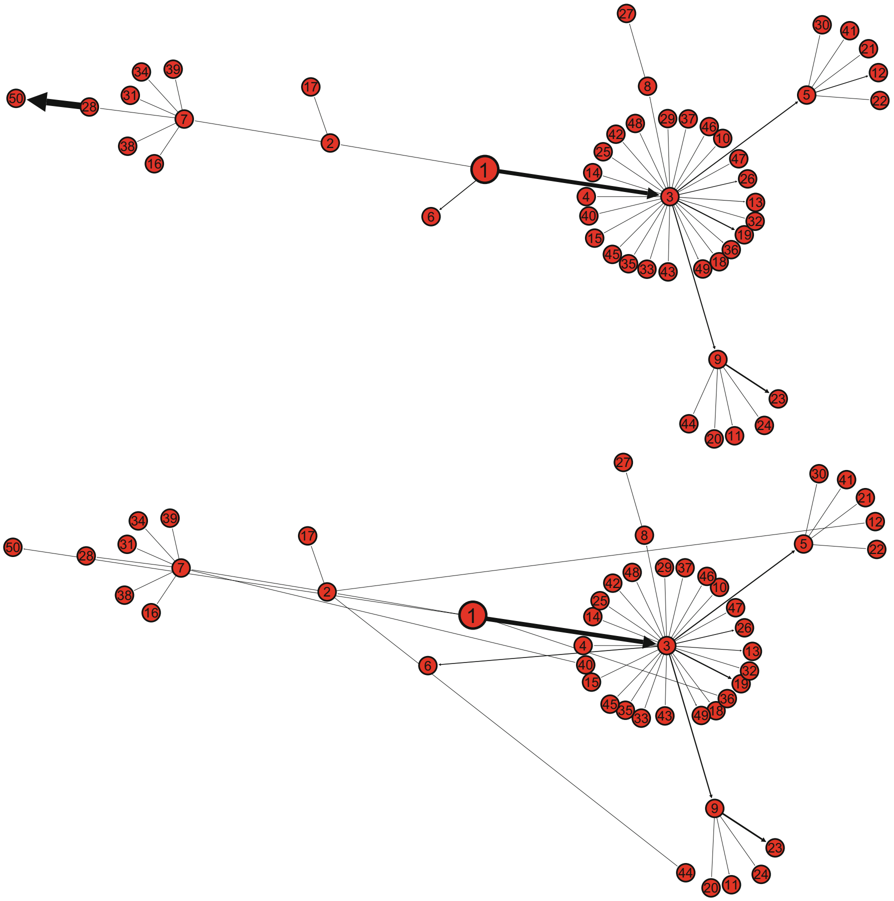

Fig. 6.9 Illustration of the actual invasion pathways (the *upper panel*) and the identified invasion pathways (the *lower panel*), during the early stage of a realization (before the appearance of 50 infected subpopulations) on the AAN. Subpopulation 1 is the seed (Sun Valley Airport in Bullhead City, Arizona) (From Wang et al. [35])

We finally examine the identifiability of an invasion case. Figure 6.10 shows the entropy and identifiability of wrongly identified $mI \mapsto S$ of 20 independent realizations on the AAN. The smaller the identifiability of an invasion case is, the more prone it is to be wrongly identified. The identifiability depicts the wrongly identified $mI \mapsto S$ more reasonably than the likelihoods entropy. The frequency of identifiability of INCs descends obviously, but that of the likelihood entropy of INCs does not clearly ascend. This statistical result indicates that the identifiability Π has a better performance to distinguish whether an invasion case is difficult to identify or not than the distinction performance of the likelihood entropy, and also tells that why some invasion cases are easy to identify, whose Π are more than 0.5,

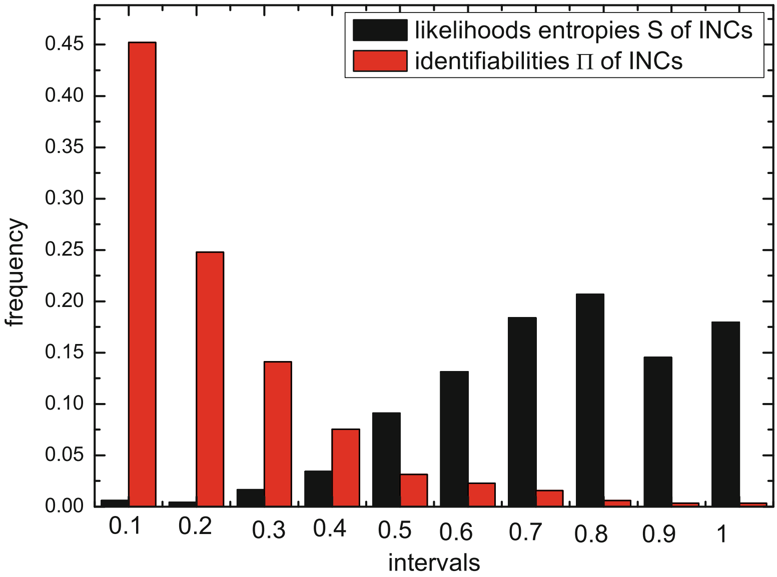

Fig. 6.10 Statistical analysis of the likelihoods entropy and identifiability of wrongly identified $mI \mapsto S$ in 20 realizations of epidemic spreading on the AAN (From Wang et al. [35])

and why some invasion cases are difficult to identify, whose Π are much less than 0.5. Here 0.5 is an empirical value.

6.5 Predicting the Epidemic Transmission

As the final part of this chapter, we now move a step further to predict the early stage of an epidemic transmission. Suppose the epidemic process starts from the patient 0 subpopulation. This subpopulation invades and infects its neighbours, and the cascading transmission proceeds. At the early epidemic stage, the time series of the number of infected individuals in each subpopulation $I_i(t)$ (i.e., the infection data) is recorded. Assume the topology of the meta-population network (including population sizes and diffusion rates, as Sect. 6.4) and the time series of the recorded infection data $I_i(t)$ until time t are available, and the focus of interest in this section is to predict which subpopulations will be infected at time step $t+1$. We consider the SI model with the diffusion of new increment of infected individuals after reaction (see Sect. 6.2 Eq. (6.3)).

6.5.1 A Prediction Algorithm

The growth of infected individuals in an infected subpopulation is governed by the infected rate β, while the diffusion process is ruled by the parameters of multinomial distribution. We first identify the infection rate β by using the method in Sect. 6.3.2, then estimate the increment $\Delta_R I_i(t)$ of $I_i(t)$ of subpopulation i after the reaction from t to $t+1$. Statistically, $\langle \Delta_R I_i(t) \rangle = \beta I_i(t) S_i(t)/N_i(t)$. To keep the population balance of each subpopulation, we assume $\langle w_{ij} \rangle = \langle w_{ji} \rangle$, i.e., $\langle N_i(t) p_{ij} \rangle = \langle N_j(t) p_{ji} \rangle$, where w_{ij} is the number of individuals that have moved from subpopulation i to subpopulation j in a unit time (e.g., a day). Thus we have $\langle N_j(t) \rangle = \langle N_j(t+1) \rangle$. At the early stage, $N_j(t) \approx S_j(t)$, and $N_j(t)$ is included in the population information of each subpopulation of meta-population network. Therefore, we estimate $\Delta_R I_j(t)$ by $\Delta_R I_j(t) \approx \beta I_j(t)/N_j(t)$.

Next we give the algorithm predicting $n (n \geq 1)$ subpopulations infected from t to $t+1$ during the diffusion process. At time step t, all susceptible subpopulations having at least one infected neighbouring subpopulation comprise set $\mathbf{S}$. We discuss the two cases of $n = 1$ and $n > 1$ in the following, and Algorithm 3 presents the pseudocode for the prediction algorithm.

(i) $n = 1$;

In this case, there is only one susceptible subpopulation infected at time $t+1$. The likelihood $\mathscr{L}_i(t+1)$ that subpopulation i in set $\mathbf{S}$ is infected at time $t+1$ is derived as

$$\begin{aligned} \mathscr{L}_i(t+1) = & 1 - (1-p_{1i})^{I_1(t)+\Delta_R I_i(t)} (1-p_{2i})^{I_2(t)+\Delta_R I_2(t)} \\ & \ldots (1-p_{mi})^{I_m(t)+\Delta_R I_m(t)}, \end{aligned} \tag{6.28}$$

where m is the number of infected neighbouring subpopulations of i at time step t. We label infected neighbouring subpopulations of i as $1, 2, \ldots, m$.

Accordingly, the most likely infected subpopulation $\hat{v}$ is predicted as

$$\hat{v} = \arg\max_i \mathscr{L}_i \prod_{j \neq i, j \in \mathbf{S}} \overline{\mathscr{L}_j}, \tag{6.29}$$

where $\overline{\mathscr{L}_j} = 1 - \mathscr{L}_j$.

(ii) $n \geq 2$;

The most likely $n (n \geq 2)$ infected subpopulations in $\mathbf{S}$ can be predicted as

$$\hat{v}_n = \arg\max_{i_k} \mathscr{L}_{i_1} \mathscr{L}_{i_2} \ldots \mathscr{L}_{i_n} \prod_{j \neq i_k, k=1,2,\ldots,n, j \in \mathbf{S}} \overline{\mathscr{L}_j}. \tag{6.30}$$

Algorithm 3 Prediction Algorithm

```
Inputs: time series of infection data $I_i(t)$ and topology of network $G$
Estimate the infection rate $\beta$
for each time step $t$
    find all possible candidate subpopulations (set $\mathbf{S}$)
    compute the likelihood $\mathscr{L}_i(t+1)$ of each subpopulation $i \in \mathbf{S}$
    rank all subpopulations $i$ by their likelihoods $\mathscr{L}_i(t+1)$
end for
Choose the subpopulation $i$ corresponding to the maximal likelihood $\mathscr{L}_i(t+1)$ as the most likely infected $i$ in the next time step
```

Note that the above method only presents the most likely infected subpopulations at the next time step. Generally, the number of possibly infected subpopulations increases sharply during the epidemic dynamics. In this case, the likelihood of the most likely infected subpopulation may be very small. Therefore, we shall rank the likelihoods and investigate the top ranking subpopulations, which help us to judge which subpopulations are prone to be infected. Let $\mathscr{P}_i = \mathscr{L}_i \prod_{j \neq i, j \in \mathbf{S}} \overline{\mathscr{L}_j}$ in Eq. (6.29). We define the infected likelihood vector $\{\mathscr{P}_1, \mathscr{P}_2, \ldots, \mathscr{P}_Z\}$ of all Z candidate subpopulations in set $\mathbf{S}$, where $\mathscr{P}_i$ is the likelihood the susceptible subpopulation i gets infected in the next time step as Eq. (6.29), $i = 1, 2, \ldots, Z$. Then we define the infected likelihood entropy $\mathscr{E}$ as

$$\mathscr{E} = -\frac{1}{\log M} \sum_{i=1}^{M} \mathscr{P}_i \log \mathscr{P}_i. \tag{6.31}$$

This entropy tells the extent of prediction difficulty at each time step. The smaller $\mathscr{E}$, the easier the prediction.

6.5.2 *Examples*

This time we select an artificial meta-population network as the simulation example of spatial epidemic prediction. We generate a scale-free network with the BA model [6], then design the diffusion rate of each edge. Note that empirically the diffusion rates [44] of air transportation networks depend on the degree of the nodes. We define the diffusion rate from node i to node j as

$$p_{ij} = \frac{b_{ij} k_j^{\hat{\theta}}}{\sum_l b_{il} k_l^{\hat{\theta}}} C, \tag{6.32}$$

where b_{ij} stands for the elements of the adjacency matrix ($b_{ij} = 1$ if i connects to j, and $b_{ij} = 0$ otherwise), C is a constant (C is assumed as available, and set as 0.005), and $\hat{\theta}$ is a parameter. We assume that parameter θ follows the Gaussian distribution

$\theta \sim N(\hat{\theta}, \delta^2) = \frac{1}{\sqrt{2\pi}\delta} exp(-\frac{(\hat{\theta}-\theta)^2}{2\delta^2})$ for each subpopulation. By setting constant C and computing the population of each subpopulation at equilibrium, the polynomial regression is employed to evaluate parameters $\hat{\theta}$ and δ^2 based on the empirical rule of $T' \sim k^{\beta'}, \beta' \simeq 1.5 \pm 0.1$, (where $T' = \sum_l w_{jl}$, and β' is approximately linear with $\hat{\theta}$ (observed in simulations). Assume $\hat{\theta} = a'\beta' + b'$, we can obtain $\hat{\theta}$, where a', b' are parameters). Therefore we can determine the diffusion rate p_{ij} along each edge. We set the whole BA meta-population network having 404 nodes (subpopulation), and fix $\langle k \rangle = 16$ ($m_0' = 9, m' = 8$) as the average degree of the BA meta-population network. The initial size of each subpopulation is $N_1 = N_2 = \cdots = N_N = 6 \times 10^5$, and the total population of the whole meta-population network is $N_{total} = 6 \times 10^5 \times 404 = 2.424 \times 10^8$.

As illustrated in Fig. 6.11, the estimation of β is close to the actual infection rate. We compare our prediction algorithm with the randomization prediction, i.e., we randomly choose a susceptible subpopulation in **S** as the most likely infected subpopulation at the next time step. Ranking distance is defined as the difference of rank of likelihood $\mathscr{L}(t+1)$ between the investigated two subpopulations i and j. In Fig. 6.12, "RankError" means the ranking distance of the corresponding infected likelihood between the predicted candidate and the actual infected subpopulation. "RandError" means the ranking distance of the corresponding infected likelihood between the randomly selected candidate and the actual infected subpopulation. As shown in Fig. 6.12, the subpopulations predicted by our algorithm are closer to the actual infected subpopulations at the next time step compared with those randomly selected subpopulations.

We further investigate why the accurate prediction of the infected subpopulation is difficult to achieve. At time step t, if any new subpopulation(s) will be infected in this realization at the next time step, $t + 1$ is called the prediction time. As shown in Fig. 6.13, we observe that the number of possible infected candidates Z increases sharply, and the infected likelihood entropy also increases (generally

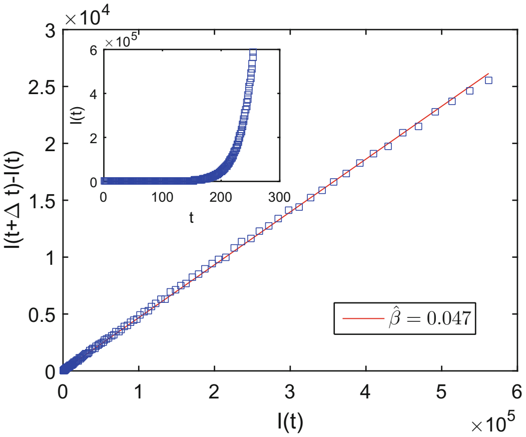

Fig. 6.11 The estimation of the infection rate β on a BA meta-population network with 404 subpopulations. The actual value of $\beta = 0.05$. Inset: The evolution of I(t) in a linear scale (From Wang et al. [36])

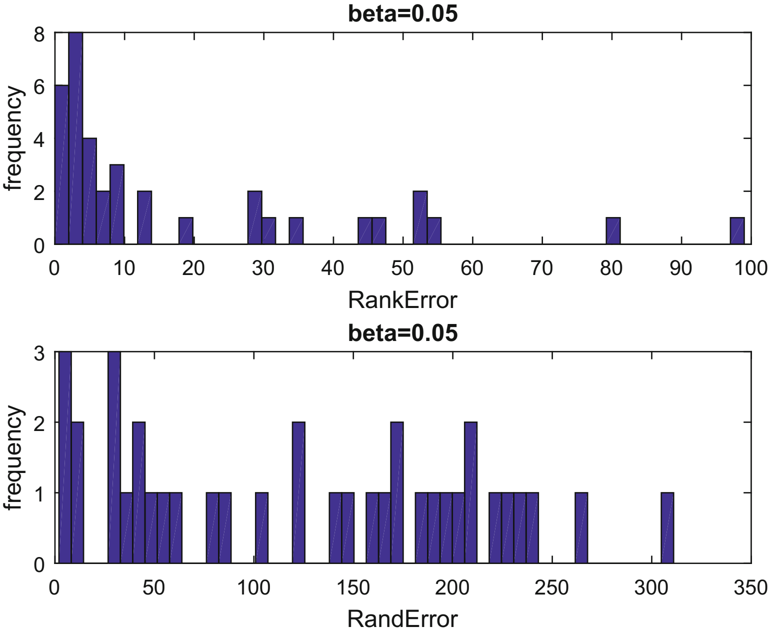

Fig. 6.12 (*Top*) The distribution of the RankError from t to $t + 1$ at the early stage of a realization (one run) of epidemic dynamics. (*Bottom*) The distribution of the RandError from t to $t + 1$ at the early stage of the same realization of epidemic dynamics. Here $t + 1$ is each time of prediction. In the realization, the infection rate $\beta = 0.05$ (From Wang et al. [36])

$\mathscr{E} > 0.5$) during the time evolution. Because the likelihoods of possibly infected subpopulations become more homogeneous as the infection prevails, indicating the infected likelihoods in the likelihood vector are not significantly different from each other, the infected likelihood entropy herein becomes large, suggesting the difficulty of accurately predicting the next infected subpopulation.

6.6 Outlook

As only a snapshot of the emergent frontier in the exciting network science, some latest advances on identification and prediction of epidemic meta-population networks have been introduced in this chapter. The future steps along this line may involve the following aspects: (1) The adaptiveness of humans deserves sufficient respect when facing the modelling, analyses and prediction of a large-scale spatial pandemic situation, and an appropriately designed role with the feedback-loop of human adaptiveness into such a complex networking system will be much appreciated. (2) The power of Big Data and cloud computing may help embed high-resolution records of human behavioural dynamics (including mobility, interaction

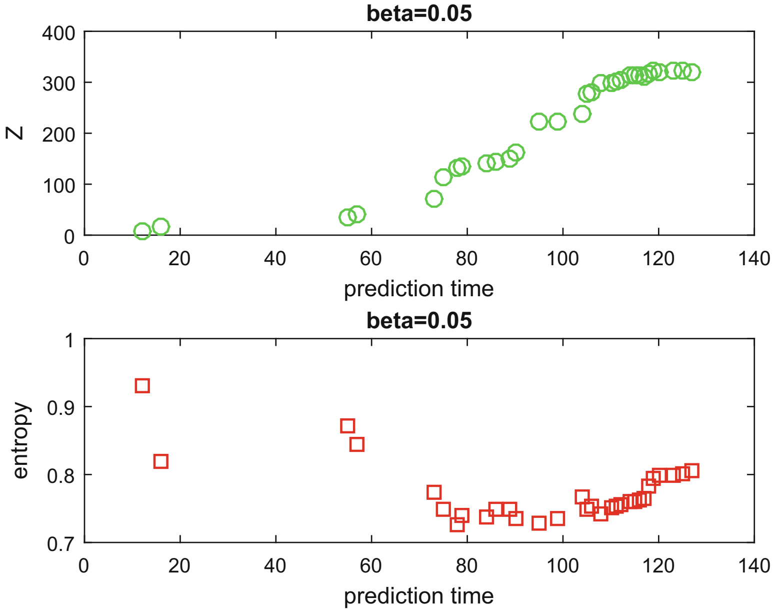

Fig. 6.13 (*Top*) The evolution of the number of possibly infected candidates Z with each prediction time. (*Bottom*) The entropy of likelihoods vector. The epidemic realization is run on a BA meta-population network with 404 subpopulations with the infection rate $\beta = 0.05$ (From Wang et al. [36])

and other non-private profiles) into the study. Nevertheless, abuse of data should be carefully avoided. (3) The verification even for the prediction of an infectious process requires the precise control means and public strategy in the viewpoints of not only mathematical results but also implementations in practice. Finally comes the end of this chapter, which may still stands at the beginning of the long journey in this exciting and challenging direction.

References

1. Wiener, N.: Cybernetics: or Control and Communication in the Animal and the Machine. MIT Press, Cambridge, MA (1961)
2. Bondy, J.A., Murty, U.S.R.: Graph Theory with Applications. Macmillan, London (1976)
3. West, D.B.: Introduction to Graph Theory. Prentice Hall, Upper Saddle River (2001)
4. Erdős, P., Rényi, A.: On the evolution of random graphs. Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17–61 (1960)
5. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. Nature **393**, 440–442 (1998)
6. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)

7. Albert, R., Barabási, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**, 47–97 (2002)
8. Wang, X., Li, X., Chen, G.: Complex Networks: Theories and Applications. Tsinghua University Press, Beijing (2006, in Chinese)
9. Newman, M.E.J.: Networks: An Introduction. Oxford University Press, New York (2010)
10. Chen, G., Wang, X., Li, X.: Introduction to Complex Networks: Models, Structures and Dynamics. Higher Education Press, Beijing (2012)
11. Keeling, M.J., Rohani, P.: Modeling Infectious Diseases in Humans and Animals. Princeton University Press, Princeton/Oxford (2008)
12. Anderson, R.M., May, R.M.: Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford (1991)
13. Heesterbeek, H., Anderson, R.M., Andreasen, V., et al.: Modeling infectious disease dynamics in the complex landscape of global health. Science **347**, aaa4339 (2015)
14. Fitch, J.P.: Engineering a global response to infectious diseases. Proc. IEEE **103**, 263–272 (2015)
15. Brockmann, D., Helbing, D.: The hidden geometry of complex, network-driven contagion phenomena. Science **342**, 1337–1342 (2013)
16. McMichael, A. J.: Globalization, climate change, and human health. N. Engl. J. Med. **368**, 1335–1343 (2013)
17. Pastor-Satorras, R., Castellano, C., Van Mieghem, P., Vespignani, A.: Epidemic processes in complex networks. Rev. Mod. Phys. **87**, 925–979 (2015)
18. Fu, X., Small, M., Chen, G.: Propagation Dynamics on Complex Networks: Models, Methods and Stability Analysis. Higher Education Press, Beijing (2014)
19. Li, X., Li, X.: A Data-driven inference algorithm for epidemic pathways using surveillance reports in 2009 outbreak of influenza A (H1N1). In: Proceedings of 51st IEEE Conference on Decision and Control (CDC), pp. 2840–2845 (2012)
20. Hufnagel, L., Brockmann, D., Geisel, T.: Forecast and control of epidemics in a globalized world. Proc. Natl. Acad. Sci. U. S. A. **101**, 15124–15129 (2004)
21. Miao, H., Xia, X., Perelson, A.S., et al.: On identifiability of nonlinear ODE models and applications in viral dynamics. SIAM Rev. **53**, 3–39 (2011)
22. Gomez-Rodriguez, M., Leskovec, J., Krause, A.: Inferring networks of diffusion and influence. In: Proceedings of 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), pp. 1019–1028 (2010)
23. Han, X., Shen, Z., Wang, W.-X., Di, Z.: Robust reconstruction of complex networks from sparse data. Phys. Rev. Lett. **114**, 028701 (2015)
24. Shah, D., Zaman, T.: Rumors in a network: who's the culprit? IEEE Trans. Inf. Theory **57**, 5163–5181 (2011)
25. Wang, Z., Dong, W., Zhang, W., Tan, C.-W.: Rumor source detection with multiple observations: fundamental limits and algorithms. In: Proceedings of the ACM Sigmetrics 2014, pp. 1–13 (2014)
26. Maeno, Y.: Discovering network behind infectious disease outbreak. Phys. A **389**, 4755–4768 (2010)
27. Eggo, R.-M., Cauchemez, S., Ferguson, N.M.: Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. J. R. Soc. Interface **8**, 233–243 (2011)
28. Wan, X., Liu, J., Cheung, W.K., Tong, T.: Inferring epidemic network topology from surveillance data. PLoS One **9**, e100661 (2014)
29. Shi, B., Liu, J., Zhou, X.-N., Yang, G.-J.: Inferring plasmodium vivax transmission networks from tempo-spatial surveillance data. PLoS Negl. Trop. Dis. **8**, e2682 (2014)
30. Yang, X., Liu, J., Zhou, X.-N., Cheung, W.-K.: Inferring disease transmission networks at a metapopulation level. Health Inf. Sci. Syst. **17**, 8 (2014)
31. Gautreau, A., Barrat, A., Barthelemy, M.: Global disease spread: statistics and estimation of arrival times. J. Theor. Biol. **251**, 509–522 (2008)
32. Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A.: Multiscale mobility networks and the spatial spreading of infectious diseases. Proc. Natl. Acad. Sci. U. S. A. **106**, 21484–21489 (2009)

33. Wang, J.-B., Cao, L., Li X.: On estimating spatial epidemic parameters of a simplified metapopulation model. In: Proceedings of 13th IFAC Symposium on Large Scale Complex Systems: Theory and Applications, pp. 383–388 (2013)
34. Wang, J.-B., Li, X., Wang, L.: Inferring spatial transmission of epidemics in networked metapopulations. In: Proceedings of 2015 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 906–909 (2015)
35. Wang, J.-B., Wang, L., Li, X.: Identifying spatial invasion of pandemics on metapopulation networks via anatomizing arrival history. IEEE Trans. Cybern. **46**, 2782–2795 (2016)
36. Wang, J.-B., Li, C., Li, X.: Predicting spatial transmission at the early stage of epidemics on a networked metapopulation. In: Proceedings of 12th IEEE International Conference on Control & Automation (ICCA), pp. 116–121 (2016)
37. Li, X., Wang, J.-B., Li, C.: Towards identifying epidemic processes with interplay between complex networks and human populations. In: Proceedings of 2016 IEEE Conference on Norbert Wiener in the 21st Century (21CW), pp. 67–71 (2016)
38. Levins, R.: Some demographic and genetic consequences of environmental heterogeneity for biological control. Bull. Entomol. Soc. Am. **15**, 237–240 (1969)
39. Rvachev, L.A., Longini, I.M.: A mathematical model for the global spread of influenza. Math. Biosci. **75**, 3–22 (1985)
40. Wang, L., Li, X.: Spatial epidemiology of networked metapopulation: an overview. Chin. Sci. Bull. **59**, 3511–3522 (2014)
41. Brooks-Pollock, E., Roberts, G.O., Keeling, M.J.: A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature **511**, 228–231 (2014)
42. Brockmann, D., Theis, F.: Money circulation, trackable items, and the emergence of universal human mobility patterns. IEEE Pervasive Comput. **7**, 28–35 (2008)
43. Wang, L., Li, X., Zhang, Y.-Q., Zhang, Y., Zhang, K.: Evolution of scaling emergence in large-scale spatial epidemic spreading. PLoS One **6**, e21197 (2011)
44. Barrat, A., Barthélemy, M., Pastor-Satorras, R., Vespignani, A.: The architecture of complex weighted networks. Proc. Natl. Acad. Sci. U. S. A. **101**, 3747–3752 (2004)