Anomalous Transport in Complex Networks

by

Christos Nicolaides

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Civil and Environmental Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

Author .....
Department of Civil and Environmental Engineering
May 7, 2011

Certified by .....
Ruben Juanes
Assistant Professor of Civil and Environmental Engineering
Thesis Supervisor

Certified by .....
Luis Cueto-Felgueroso
Postdoctoral Associate of Civil and Environmental Engineering
Thesis Supervisor

Accepted by .....
Heidi M. Nepf
Chair, Departmental Committee for Graduate Students

Anomalous Transport in Complex Networks

by

Christos Nicolaides

Submitted to the Department of Civil and Environmental Engineering
on May 7, 2011, in partial fulfillment of the
requirements for the degree of
Master of Science in Civil and Environmental Engineering

Abstract

The emergence of scaling in transport through interconnected systems is a consequence of the topological structure of the network and the physical mechanisms underlying the transport dynamics. We study transport by advection and diffusion in scale-free and Erdős–Rényi networks. Using stochastic particle simulations, we find anomalous (nonlinear) scaling of the mean square displacement with time. We show the connection with existing descriptions of anomalous transport in disordered systems, and explain the mean transport behavior from the coupled nature of particle jump lengths and transition times. Moreover, we study epidemic spreading through the air transportation network with a particle-tracking model that accounts for the spatial distribution of airports, detailed air traffic and realistic (correlated) waiting-time distributions of individual agents. We use empirical data from US air travel to constrain the model parameters and validate the model's predictions of traffic patterns. We formulate a theory that identifies the most influential spreaders from the point of view of early-time spreading behavior. We find that network topology, geography, aggregate traffic and individual mobility patterns are all essential for accurate predictions of spreading.

Thesis Supervisor: Ruben Juanes
Title: Assistant Professor of Civil and Environmental Engineering

Thesis Supervisor: Luis Cueto-Felgueroso
Title: Postdoctoral Associate of Civil and Environmental Engineering

Acknowledgments

I am grateful to my research advisors, Prof. Ruben Juanes and Dr. Luis Cueto-Felgueroso for their valuable guidance during this project; to Prof. Marta González for her useful comments and advices; to my friends & fellow students, among them Birendra Jha, Michael Szulczewski, Chris MacMinn, Ben Scandella, and many others, for their camaraderie and occasional commiseration; to my fiancée, Elli Loizidou, for her encouragement; and to my parents, Andreas and Maria Nicolaidou, and sister, Ioanna Nicolaidou, for their confidence in my abilities. Thank you!

Contents

List of Figures

Chapter 1

Introduction

Complex systems are very often organized under the form of networks with nodes and edges between them. During the last decade the idea of complex networks cause a breakthrough to the scientific community and have long been the subject of many studies in mathematics, mathematical sociology, computer science and genetics and medicine. Complex weblike structures describe a wide variety of systems of high technological and intellectual importance. For example, the cell is best described as a complex network of chemicals connected by chemical reactions; the Internet is a complex network of routers and computers linked by various physical or wireless links; fads and ideas spread on the social network, whose nodes are human beings and whose edges represent various social relationships; the World Wide Web is an enormous virtual network of Web pages connected by hyperlinks; metapopulation networks, where the nodes represent populations and the connections flows of individuals between different areas, have used to study global epidemic spreading; cellular functions are carried out by a complex network of genes, proteins, and metabolites that interact through biochemical and physical interactions; These systems represent just a few of the many examples that have recently prompted the scientific community to investigate the mechanisms that determine the topology of complex networks.

1.1 Structure Of Real World Nets

The study of complex systems began with the effort to identify their structure and develop models that can reproduce their characteristics. The first model was proposed by Erdős and Rényi [15] at the end of the 1950s and was at the basis of most studies until recently. They assumed that complex systems are wired randomly together, a hypothesis that was adopted by sociology, biology, and computer science. It had considerable predictive power, explaining for example why everybody is only six handshakes from anybody else [38] , a phenomenon observed as early as 1929 and is well known as *'the six degrees of separation'*. However, this model failed to explain the big clustering, a common property of social networks where cliques form, representing circles of friends or acquaintances in which every member knows every other member. This latter property is characteristic of ordered regular lattices.

The interest in networks was however renewed in 1998 by Watts and Strogatz [40], who extracted stylized facts from real-world networks and proposed a simple, new model of random networks. They proposed a model that interpolates between an ordered finite-dimensional lattice and a random graph. This model combines small average shortest path and big clustering, characteristics of a wide variety of social complex systems.

In the above models, the number of nodes a node is connected (node degree or connectivity) is more or less the same for all the nodes. More specifically, the degree distribution of a random graph is a Poisson distribution with a peak at $P(< k >)$, where $< k >$ is the average degree. However, one of the most interesting developments in our understanding of complex networks was the discovery that for most large networks the degree distribution significantly deviates from a Poisson distribution. In particular, for a large number of networks, including the World Wide Web [2], the Internet [16], or metabolic networks [19], the degree distribution has a power-law tail,

$$P(k) \sim k^{-\gamma}. \tag{1.1}$$

Such networks are called scale free [3]. While some networks display an exponential

tail, often the functional form of $P(k)$ still deviates significantly from the Poisson distribution expected for a random graph.

The origin of the power-law degree distribution observed in networks was first addressed by Barabási and Albert (1999) [3], who argued that the scale-free nature of real networks is rooted in two generic mechanisms shared by many real networks: Growth and Preferential Attachment: (i) *Growth* Starting with a small number (m_0) of nodes, at every time step, we add a new node with $m(\leq m_0)$ edges that link the new node to m different nodes already present in the system. (ii) *Preferential Attachment* When choosing the nodes to which the new node connects, we assume that the probability that a new node will be connected to node i is proportional to the degree k_i of node i, such that

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{1.2}$$

Numerical simulations indicate that this network evolves into a scale-invariant state with the degree of a node following a scale-free distribution with power $\gamma_{BA} = 3$.

1.2 Dynamical Processes in Complex Networks

Recent work has led to fundamental advances in our understanding of flow and transport properties of complex networks. These include the analysis of the conductance between two arbitrarily chosen nodes in scale-free or Erdős and Rényi networks [25]—an analysis that has been extended to the case of multiple sources and sinks [10, 9], and to weighted networks [23]. The dynamics of spreading through scale-free networks have been studied by means of *diffusive* random walks [32, 17, 8, 18, 5], which have led to scalings for the mean first passage time (MFPT) in terms of network centrality [32] and modularity [18]. Studies to date leave open, however, the question of how *directional bias* impacts transport. Bias, or drift, occurs naturally in many network systems, including advection from a flow potential [25], agent spreading in topologies with sources and sinks, such as utility networks [11] and freely diffusing molecules in

tissue [41], and tracer diffusion in suspensions of swimming microorganisms [22].

Chapter 2

Potential Driven Diffusion in Spatial Embedded Complex Networks

Here, we study the scaling properties of physical transport in scale-free and Erdős–Rényi networks. The fundamental question is whether the topology of the network, the underlying physical mechanisms, or both, control the scaling properties of transport in complex networks. Transport is physical in the sense that, as observed in most natural settings, the spreading process is conservative and driven by the interplay between advection—which derives from a flow potential—and diffusion, which we model through random walks. We show that advection leads to anomalous (non-Fickian) transport, as evidenced by the nonlinear time scaling of the mean square displacement (MSD) of tracer particles migrating through the network. The simulation results suggest that a mean-field theory such as Continuous Time Random Walk (CTRW), which describes transport from a joint probability distribution of particle jump lengths and transition times [30, 35, 21, 28], may be used to capture the average transport behavior. We show that coupling between space and time [7, 14] is essential to describe advective transport in a complex network.

2.1 Physical Setting

To study the scaling properties of advection-driven transport, we construct scale-free networks, characterized by a power law connectivity distribution, $P(k) \sim k^{-\gamma}$, where k is the number of links attached to a node, and γ is the characteristic exponent of the network, with $2 \leq \gamma \leq 4.8$. We generate graphs following the Molloy–Reed scheme [29]. Given the size of the network (N nodes), the degree exponent γ, and fixing the minimum degree to be $k_{\min} = 2$, we use the relations from Aiello *et al.* [1] to obtain the degree sequence from a power law distribution. We then produce a list of k_i copies of each node i, and attach links between nodes by randomly pairing up elements of this list until none remain. We disallow double links, as well as self-linked nodes. This algorithm generates networks of excellent accuracy for arbitrary exponents γ, even for relatively small network sizes. We generate an Erdős–Rényi network by attaching a link to each pair of nodes with probability $p = 0.01$. We assume for simplicity that the nodes in our model network are uniformly spaced in the unit square (Fig. 2-1). The positioning of nodes on a square lattice is an idealization of networks with spatial embedding that exhibit power-law connectivity, including power and distribution networks [11], and transportation and biological systems [39]. Networks embedded in metric spaces often exhibit a decay of nodal connectivity with distance [34, 24]—we have not included this distance-dependent connectivity in our study which therefore allows for long-ranged links.

2.2 Potential-Driven Transport Process

Flow through the network is driven by a scalar (potential) field, and satisfies conservation of mass. At every node i we impose $\sum_j u_{ij} = 0$, where u_{ij} is the flux through the link connecting nodes i and j. The fluxes are given by

$$u_{ij} = -\lambda_{ij} \frac{P_i - P_j}{d_{ij}}, \tag{2.1}$$

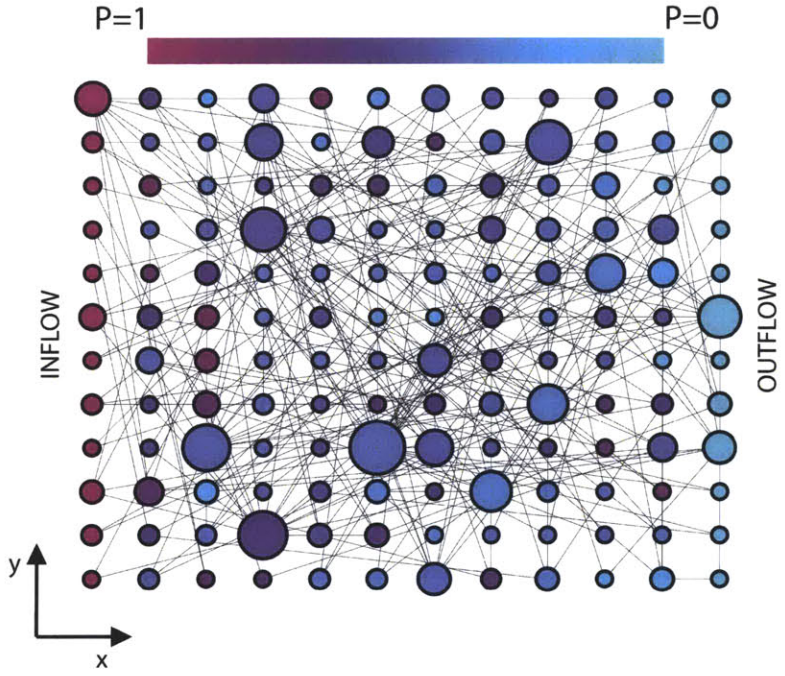

Figure 2-1: Schematic representation of our model network [31]. Nodes are distributed uniformly on the unit square. Their connectivity (represented by node diameter) follows a power law distribution. Flow through the network derives from a potential P that varies between 1 (left boundary) and 0 (right boundary).

where P_i and P_j are the flow potentials at nodes i and j, respectively, and d_{ij} is the Euclidean distance between the two nodes. This problem is exactly analogous to the electric conductance model of [25, 10, 9, 23], except that in our model the relation between velocity and potential difference is modulated by the inverse of link length. Here, we assume that the conductivity is the same for all links, and takes a value $\lambda_{ij} = 1$. To elucidate the essential mechanisms governing transport, we study a simple setting of flow from left to right by fixing the potential $P = 1$ at the inflow nodes (left boundary), and $P = 0$ at the outflow nodes (right boundary). Flow is confined between the top and bottom boundaries. Inserting the flux equation [Eq. (2.1)] into the mass conservation condition at each node, and imposing the boundary conditions, leads to a linear system of algebraic equations to be solved for the flow potential at the nodes, $\{P_j\}$, which are stationary and do not evolve in time. This system of equations is solved with a direct solution method such as Gaussian elimination.

For different values of the degree exponent γ, the distributions of nodal potential collapse in the range $P \in (0.4, 0.6)$ [Fig. 2-2(a)]. Within this range, the flow potential seems to be normally distributed (dashed line). The velocity at the links, u, follows a distribution that exhibits a power law tail [Fig. 2-2(b)]. In the range $\gamma \in [2, 3.2]$, the exponent of the velocity power law distribution is well approximated as $\nu \approx \gamma + 1$ [Fig. 2-2(b), inset]. As a consequence of this behavior, a slower decay in network connectivity (smaller γ) results in increased probability of observing large velocities at the links. Interestingly, the flow potential distribution for a random (Erdős–Rényi) network has a very different shape from that of scale-free networks, yet it also leads to a power law tail distribution of the link velocities. Note that for a regular lattice, the distribution of flow potentials is uniform, and the distribution of velocities is a double Dirac delta function.

2.3 Random Walks

We investigate the scaling properties of transport in scale-free networks through stochastic particle simulations. This allows us to explore the transition between the

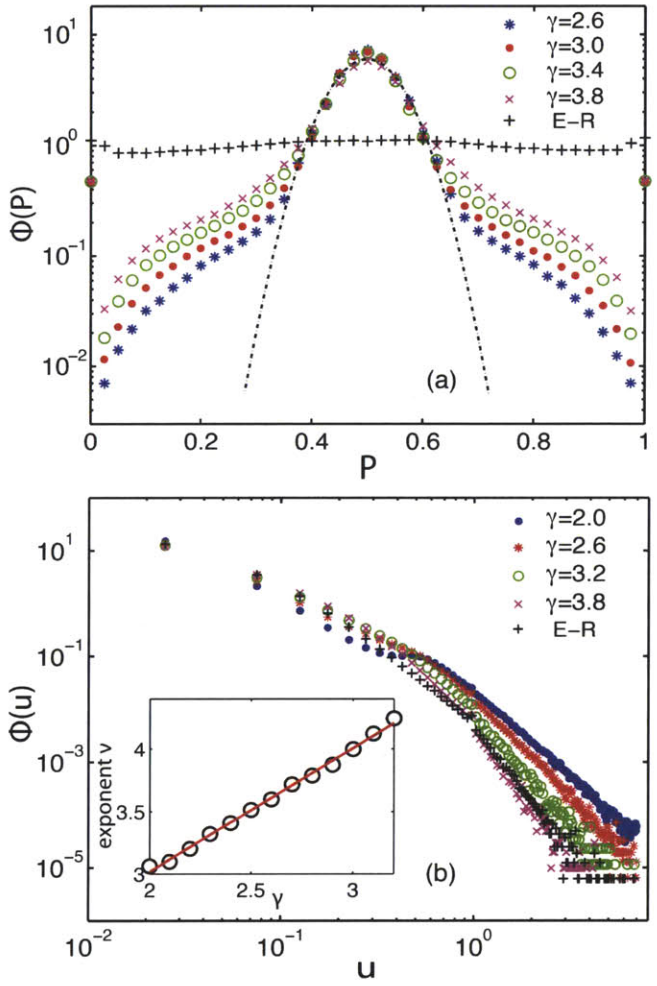

Figure 2-2: (a) Pressure distributions, $\Phi(P)$, for scale-free networks with different degree exponents, γ, as well as for an Erdős–Rényi network (E-R). The dashed line is a Gaussian fit with mean $\bar{P} = 0.5$ and standard deviation $\sigma_P \approx 0.08$. We use networks of size $N = 8100$ nodes, and the results are averaged over 250 realizations. (b) Velocity distributions, $\Phi(u)$, for different network types. Inset: Exponent ν of the velocity distribution power-law tail plotted against the connectivity exponent $\gamma \in [2, 3.2]$; shown also is a least-squares linear interpolation $\nu = (0.99 \pm 0.02)\gamma + (1.03 \pm 0.05)$.

purely diffusive and advective regimes. Particles move along the links of the network according to the local advective velocity field obtained from the steady-state potential solution, together with an additional random diffusive component sampled from a Gaussian distribution:

$$X_{N+1} = X_N + \delta X_N, \tag{2.2}$$

$$\delta X_N = u_{ij}\delta t + \sqrt{2D\delta t}\,\xi, \tag{2.3}$$

$$X_{N+1} \geq 0, \tag{2.4}$$

where X denotes the coordinate along the link, D is the diffusion coefficient, and $\xi = N(0,1)$. At $t = 0$, we inject particles at random at the inflow nodes. A particle located at node i at a given time step chooses one of the *outgoing* links to "walk" on it. Link selection is performed at random, with probability that is linearly proportional to the magnitude of the link velocity. The term *outgoing* refers here to links with velocity vectors pointing from the initial position of the particle (node i), to the destination connected nodes. Constraint (2.4) forces particles walking on a link ij to stay on it until they reach the destination node j. Particles are removed when they reach the outflow nodes.

The key parameter in this stochastic process is the ratio between advective and diffusive jump lengths,

$$A = \frac{\bar{u}\delta t}{\sqrt{2D\delta t}}, \tag{2.5}$$

where $\bar{u}$ is the characteristic velocity at the links. This computational quantity plays the role of the Péclet number ($Pe = \bar{u}\bar{d}/D$) in physical fluid flow, where $\bar{d}$ is the characteristic length of the links. The purely advective and diffusive limits correspond to $A \to \infty$ and $A = 0$, respectively. A second computational parameter is $\varepsilon = \bar{u}\delta t/(A\bar{d})$, which measures the effect of the boundary constraint (2.4). By fixing the value of ε, we impose a constant ratio between the number of advective jumps affected by the constraint, and the average number of jumps needed for a particle to move from one node to another. This permits a rigorous interpretation of the results from

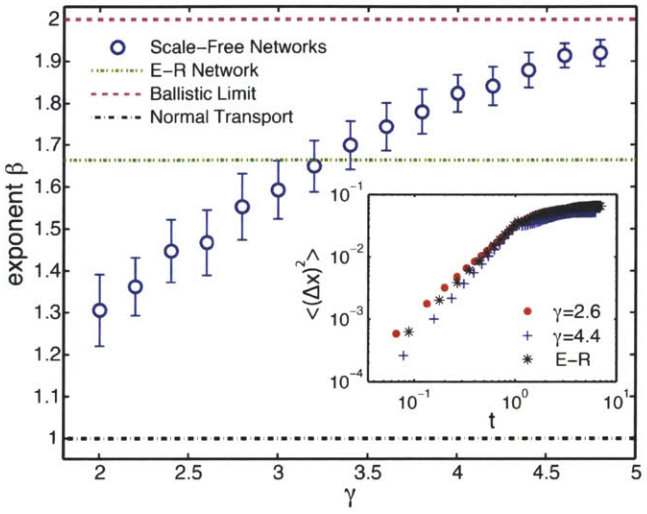

Figure 2-3: The particle MSD follows a power law in time during early times, before it is influenced by boundary effects (inset). The main plot shows the scaling exponent β of the particle MSD, for advection-dominated transport ($A \gg 1$) in scale-free networks, plotted against the connectivity degree exponent γ. Transport is anomalous, and falls in the superdiffusive regime ($\beta > 1$). Shown also are the values of β for an Erdős–Rényi (E-R) network—for which transport is also superdiffusive—and a regular lattice–for which transport is normal, $\beta = 1$. We use networks with $N = 8100$ nodes, and set $A = 7$ and $\varepsilon = 0.0014$. We use 3000 particles and we average over 100 realizations. Within each realization, we construct the network, solve the flow potential equation, and advance particles by advection and diffusion, according to the flow field from the potential solution.

particle tracking simulations. We analyze the Euclidean mean square displacement (MSD, also known as dispersion, or variance of the particle plume) in the direction of the mean flow: $\langle \Delta x^2(t) \rangle = \langle x^2(t) \rangle - \langle x(t) \rangle^2$, where $\langle \cdot \rangle$ denotes averaging over all the particles.

In the advective limit ($A \gg 1, \varepsilon \ll 1$), the MSD follows a power law at early times, $\langle \Delta x^2(t) \rangle \sim t^\beta$, and then saturates due to finite size effects (Fig. 2-3, inset). We compute the scaling exponent β in the pre-asymptotic regime for scale-free networks with different exponents γ, as well as for an Erdős–Rényi (E-R) network. Our simulations show that advection-dominated transport on scale-free networks is anomalous, and exhibits superdiffusive behavior, $\beta > 1$ (Fig. 2-3). The exponent β increases monotonically with γ. For large γ, it asymptotically reaches the ballistic limit, $\beta = 2$. We have confirmed that our results are independent of network size for $N > 4000$ (Fig. 2-4).

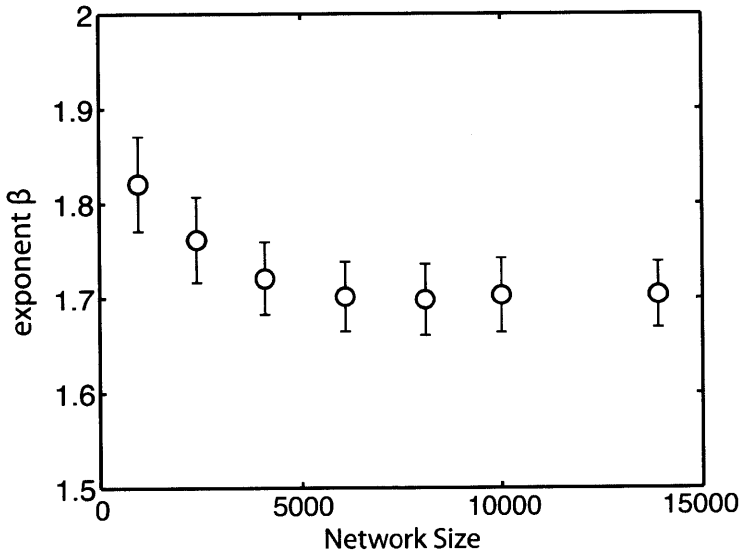

Figure 2-4: The scaling exponent β as a function of the network size for scale-free network with connectivity exponent $\gamma = 2.6$. We use 3000 particles, parameters value $A = 7$ and $\varepsilon = 0.0014$, and we average over 100 realizations

The scaling properties of transport in scale-free networks depend on the relative dominance between advective and diffusive mechanisms (Fig. 2-5). We perform simulations varying the ratio between the advective and diffusive jump lengths by

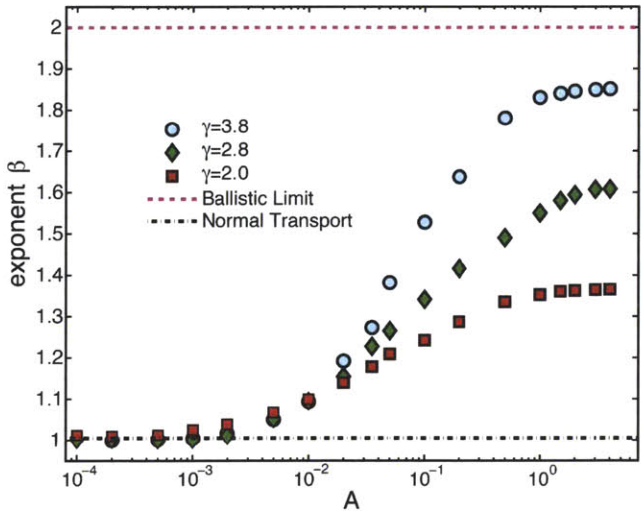

Figure 2-5: Purely diffusive transport ($A \ll 1$) in scale-free networks leads to linear scaling of the mean square displacement ($\beta \approx 1$). As A increases, we observe a transition from normal transport to anomalous transport, with scaling exponents, β, that reach different asymptotic values in the superdiffusive regime, depending on the network topology (Fig. 2-3). We use networks with $N = 8100$ nodes, fixed $\varepsilon = 0.04$, and 3000 particles. Simulation results are averaged over 100 realizations.

controlling the value of parameter A. For small values of A, transport is governed by diffusion, while for large values advection dominates. The scaling exponent of the MSD increases monotonically with A. This implies a transition from the normal scaling ($\beta = 1$) of purely diffusive transport, to anomalous scaling ($\beta > 1$) in the advection-dominated limit. The asymptotic value of β as $A \to \infty$ is controlled by the topology of the network (Fig. 2-3).

2.4 The Continuous Time Random Walk Framework

The observed scaling can be understood within the Continuous Time Random Walk (CTRW) framework [30, 35, 6, 20]. This model of transport assumes that the statistics of particle spreading can be fully characterized by a joint probability distribution, $\psi(\chi, \tau)$, of particle jump lengths, χ, and waiting time between jumps, τ, whose

marginal distributions $\lambda(\chi)$ and $w(\tau)$, respectively, can be derived by integration. It is well understood that broad distributions of jump lengths or transition times may lead to anomalous transport [30, 35, 6, 20]. To test the ability of CTRW to describe transport in scale-free networks, we perform 1D random walks by sampling the empirical joint distributions $\psi(\chi, \tau)$ recorded from our network simulations. In the advection limit, the jump length can be represented by the length of the link, d. Since the network is restricted to a finite domain, d is bounded and is found to be well approximated by a Kumaraswamy distribution [Fig. 2-6]. In contrast—and due

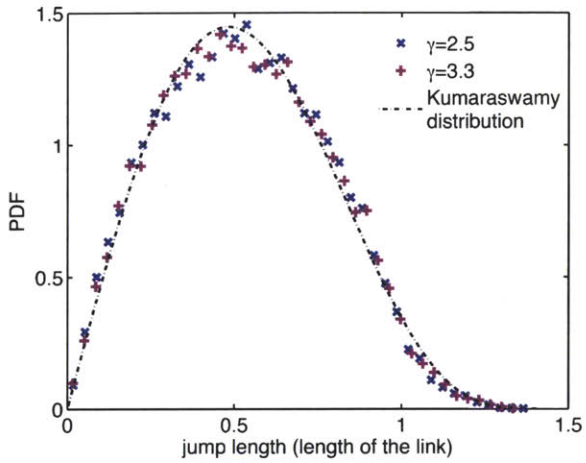

Figure 2-6: Distribution of link lengths in our embedded Erdős–Rényi and scale-free networks for different values of the degree exponent, $\gamma = 2.5$ and 3.3. This distribution of link lengths is well approximated by a double-bounded Kumaraswamy distribution $\Phi(d) \sim (d/\sqrt{2})(1 - (d/\sqrt{2})^2)^4$.

to the broad-ranged velocity distributions—the transition times between connected nodes, $\tau = d/u$, exhibit a wide range distribution with power law tail, $w(\tau) \sim \tau^{-(1+\delta)}$ [Fig. 2-7(a)], where $\delta \approx 1.9$ for all degree exponents γ. In the case of decoupled CTRW [$\psi(\chi, \tau) = \lambda(\chi)w(\tau)$] in an infinite lattice under a bias, we would expect an asymptotic late-time scaling of the MSD, $\langle \Delta x^2(t) \rangle \sim t^{1.1}$ [35, 36, 28], in contrast with the observed scalings (Fig. 2-3) The lack of quantitative agreement between decoupled CTRW and the scaling observed in our network simulations highlights the importance of the coupling between jump length and waiting time [Fig. 2-7(a), insets] in the MSD, especially at early times [Fig. 2-7(b), bottom-right inset]. A 1D coupled

CTRW qualitatively recovers the monotonic increasing trend of the scaling exponent β as a function of γ [Fig. 2-7(b)]. Quantitative agreement requires introducing additional information from the network simulations, such as the possibility of back-flow, and linear interpolation of the particle positions between jumps [Fig. 2-7(b), top-left inset]. The transition to normal transport for small γ (Fig. 2-3) is consistent with the convergence of the joint distribution of jump lengths and transition times towards the geometric limit $\tau \sim d^2$ [Fig. 2-7(a), top-left inset]. This unexpected diffusion-like scaling, whose origin appears to lie in the existence of very well connected nodes, is responsible for bounding anomalous transport behavior.

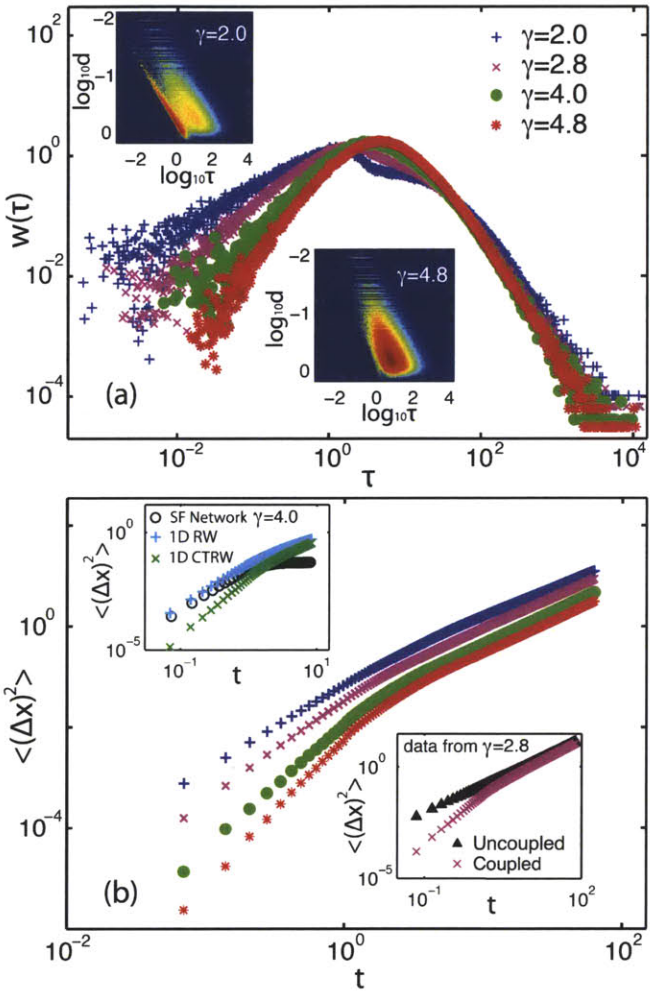

Figure 2-7: (a) Marginal distribution of waiting times for advection-dominated transport and different scale-free networks. Insets: Joint distribution $\psi(d, \tau)$ for two different scale-free networks ($\gamma = 2.0$ and 4.8). Link length and transition time exhibit strong coupling for $\gamma = 2.0$, for which the joint distribution collapses around the scaling $\tau \sim d^2$. The results are for network size $N = 8100$ nodes and over 300 realizations. (b) MSD for coupled and unidirectional CTRW, using data from different network types. Bottom-right inset: MSD for coupled and uncoupled biased, 1D CTRW. Top-left inset: MSD from network simulations compared with the results from 1D coupled CTRW, and 1D RW with back-flow and linear interpolation along the jumps. By *back-flow* we mean that the velocity along links in the network simulations is predominantly in the positive x-direction, but the flow is 'backwards' for some links. The 1D random walk simulations with backflow are based on taking a fraction of the jumps in the negative x-direction, to better approximate the actual 2D network simulations. In the classical CTRW framework, particles are assumed to 'wait' at a given location until they jump to a new one. By *interpolation* we mean that we post-compute the mean square displacement (MSD) at a given time as if the particles had experienced a constant velocity between jumps—a behavior that more closely resembles the setting of the 2D particle tracking network simulations.

Chapter 3

Influential Spreaders in Air Transportation Network

The understanding of explosive spreading of infectious diseases is crucial to global population health, especially given the growing connectivity of an increasingly vulnerable human mobility system. In this chapter, we build an agent mobility model using single-agent modeling strategy in which each individual is tracked in time. The system evolves following a stochastic microscopic dynamics and at each time step it is possible to record average quantities such as the density or the geographical spreading of infected agents. Each individual particle has its own itinerary based on real data, latest advances in understanding human mobility patterns and on a Continuous Time Random Walk framework with correlated waiting times. Individual mobility patterns are essential to capture the early time spatial characteristics of an epidemic outbreak.

3.1 Available Data

We have available data that count for flights from all airlines with at least one origin or destination inside US (including Alaska) for the period of time between January 2007 to July 2010 provided by Federal Aviation Administration [1]. The data divided in two datasets: A) T-100 Segment and B) T-100 Market. In Market, a passenger

[1] http://www.faa.gov/

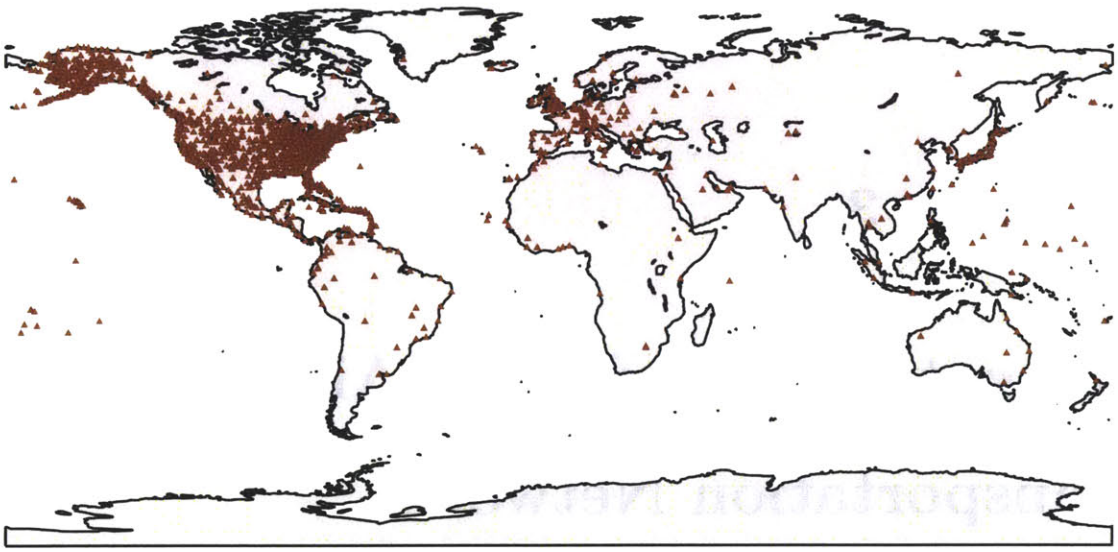

Figure 3-1: World map with the location of the airports in the US database from the Federal Aviation Administration.

is "enplaned" and is counted only once as long as he/she remains on the *same flight number* up to the final destination. In segment, a passenger is "transported" and is counted for each leg of the trip. Hence, from the Market dataset we can extract effective flows (not everything) from one airport to the other (W matrix), while from the segment database we can calculate exact traffic in each segment of the network (W^{Tr} matrix). We also have available itinerary data from a major US airline, that account for tickets have been bought for domestic travels during the period July 1^{st} and October 31^{st}, 2004 [2].

3.2 Agent Mobility Model

The Air Transportation network consist of V Airports and E edges. The network characterized by the weighted matrix W where its entry w_{ij} gives the number of passengers travel from the airport i with final destination j. To each airport j we assign a population given by the empirical relation $N_j \sim T_j^{\alpha}$, where $\alpha = 0.5$ and $T_j = \sum_j w_{ij}$ is the total traffic of the airport j [12].

[2]Barnhart group (Transportation@MIT)

Each individual at a given airport choose a destination airport to travel [27] according to the effective fluxes W with probability:

$$p_{ij} \sim w_{ij} \tag{3.1}$$

Based on the fact that W accounts only for travels where the individuals remain under the same flight number, i.e. does not reflect actual flows, we set a constant small probability for the agent to choose any other destination.

The individual travel to the final destination airport by choosing the route with the less total cost, ie the route that minimize

$$C = \sum c_{i,j} \sim \sum \frac{d_{ij}^{0.2\pm0.1}}{(w_{ij}^{Tr})^{0.3\pm0.2}}. \tag{3.2}$$

Where the sum is over all the segments (i, j) of the trip. This formula combines the need for small number of connection flights to the destination, routes with total large traffic and intermediate steps, in a sense somewhere between origin and destination. The noise I put there acounts for the different social-economic situation of each agent: rich people care only about minimize number of connections however poor people care about the money cost that is highly (negative) correlated with the flux in the links. The above model based on the currently explosion of usage websites that give many alternative choices of flight between an origin and a destination (Expedia, Orbitz etc).

If the route is with intermediate stops, at the connecting airport, the agent will wait for a short time period chosen by real data of scheduled waiting times at connecting airports (Fig. 3-2) In general, when an agent is off ground, flights with the average plane velocity of $u = 650km/h$.

After the agent arrives to the destination, he/she choose one waiting time from a waiting time distribution (Fig. 3-2) and stays there. The way we assign populations in each airport as a function of total traffic, allow us to use same waiting time distribution for all agents.

When the waiting time elapse, the agent use the *Back to Home* rule: 98.5% of the agent itineraries include only one final destination and return home (i.e. at the loca-

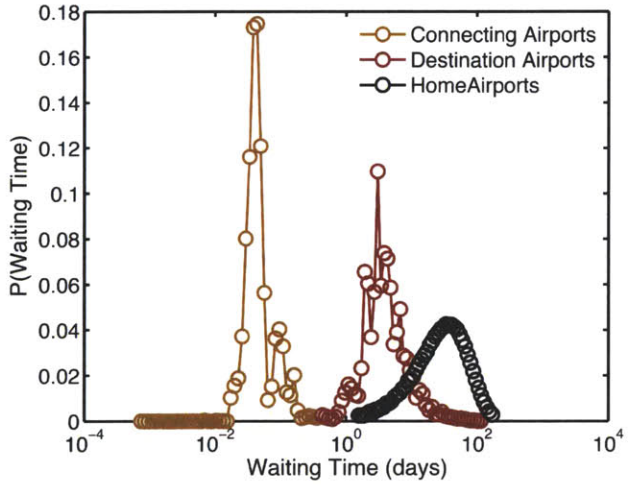

Figure 3-2: Waiting time distribution at the connecting airports [4], at the final travel destinations (individual Ticket data from a major US airline between July 1st and September 30th, 2004 for domestic travels) and at the agent's home (artificial). We make use of the fact that the time spent an agent home is order of magnitude larger than the time spent at a travel destination

tion the agent assigned initially). 1.05% of the itineraries include 2 final destnation and 0.45% include 3 final destinations. For the above rule we use the individual ticket data ($\sim$ 2.5 million round trip itineraries) from a major US airline between July 01, 2004 and September 30, 2004 for domestic flights.

If the agent is at "home", or continue his travel, he choose one destination airport using two recently applied rules for explaining human mobility patterns: Exploration and Preferential Visit [37]:

Exploration: the probability to visit a *new* airport is:

$$P_{New} = \rho S^{-\gamma} \tag{3.3}$$

where S is the number of unique airports an agent has visited in the past. We will use $\gamma = 0.21 \pm 0.02$ and ρ ($\rho \geq 0$) from a Gaussian distribution with mean $\overline{\rho} = 0.6$ that fit the mobile phone data of human mobility [37].

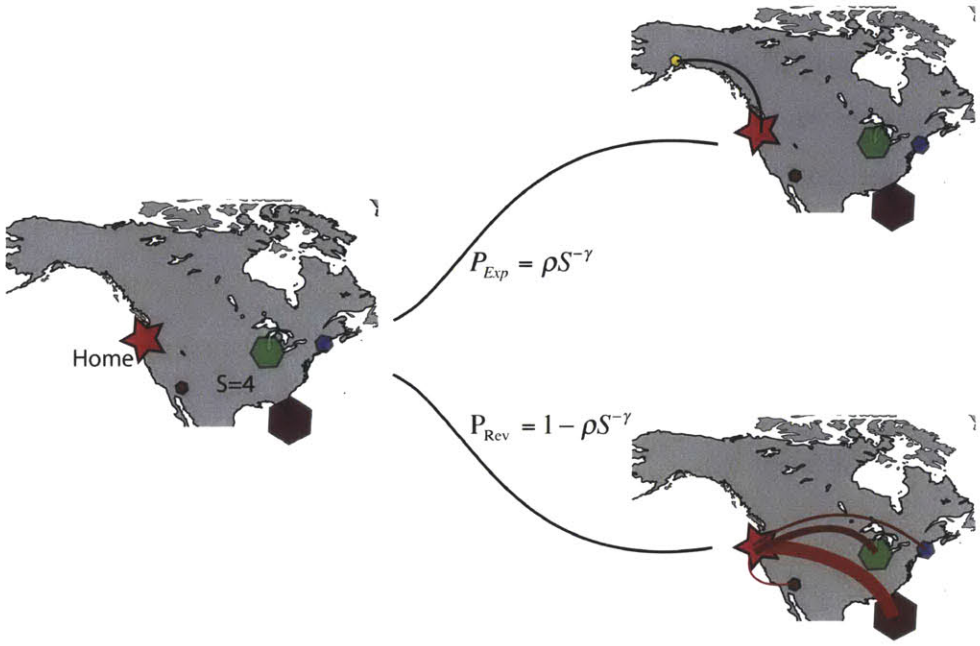

Figure 3-3: Exploration and Preferential Visit. With star is represented the home of the individual (Seattle) and with the hexagons, the already visited places. The size of the hexagon represents how many times is visited by the agent in the past.

Preferential visit: With complementary probability

$$P_{Ret} = 1 - \rho S^{-\gamma} \tag{3.4}$$

the agent will visit airports that already have visited. In this case the probability Π_i to visit an airport i is chosen proportional to the number of previously visits at location i, f_i

$$\Pi_i = f_i \tag{3.5}$$

We allow the agent mobility model to advance for a period of time ($\sim$ 1 year) in order to 'train' the agents and avoid early time instability that the exploration and preferential visit rules may cause.

3.3 Infection Model

We use the compartmental SIR (Susceptible-Infected-Recovered) and SIS (Susceptible-Infected-Susceptible) models to study epidemic spreading in our mobility model. The SIR model is characterized by a set of reaction equations:

$$S + I \xrightarrow{\beta} 2I$$

$$I \xrightarrow{\mu} R, \tag{3.6}$$

while the SIS can be described by:

$$S + I \xrightarrow{\beta} 2I$$

$$I \xrightarrow{\mu} S \tag{3.7}$$

The compartmental models above conserves the number of individuals and are characterized by the reproductive number $R = \beta/\mu$, which determines the average number of infectious individuals generated by one infected individual in a fully susceptible population. The epidemic is able to generate a number of infected individuals larger that those who recovered only if $R > 1$, yielding the classic result for the epidemic threshold [33]. On the other hand, if the spreading rate is not large enough to allow a reproductive number larger than one (that is $\beta > \mu$), the epidemic outbreak will affect only a negligible portion of the population and will die out in a finite amount of time. This result is valid at the level of each subpopulation (airport) and holds also at the metapopulation level (network of airports) where $R > 1$ is a necessary condition to have the growth of the epidemic [13].

Under the assumption of homogeneous mixing within a city, contacts between Susceptible and Infectious individuals are uniformly distributed. Let β be the contact rate for each time step, defined as the average number of individuals with whom an infectious individual will make sufficient conduct to pass the infection in the time interval $(t, t + \delta t)$. The average number of new infections caused by one infectious

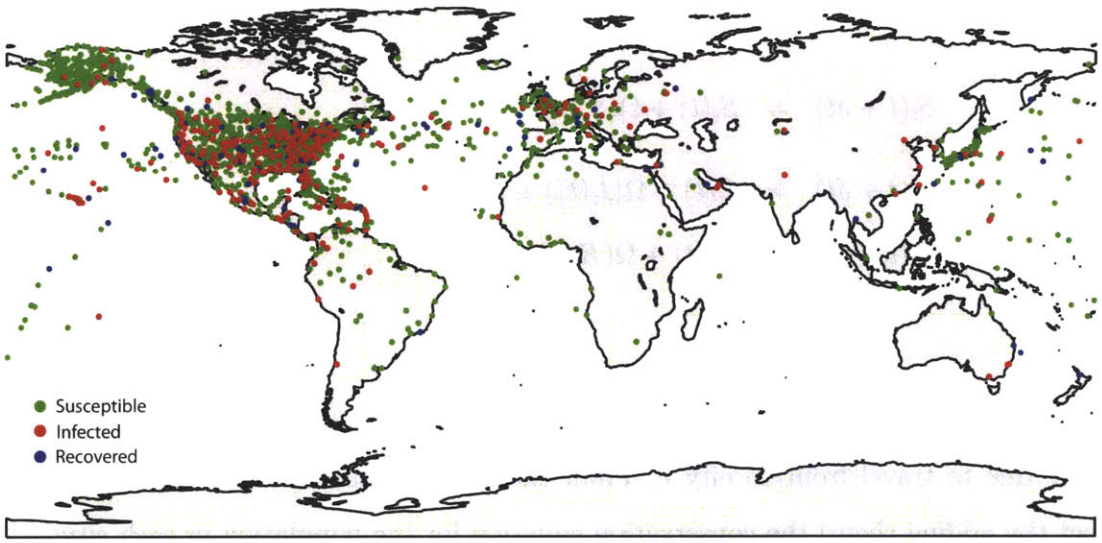

Figure 3-4: Snapshot of susceptible, infected and recovered agents during airline traffic-driven disease spreading from Los Angeles Int. Airport (LAX), after ~10 days. In this particular simulation, we used an SIR model with parameters β=0.01 (infection rate) and μ=0.004 (recovery rate), and a total of 50,000 agents.

person in that city on that time $(t, t + dt)$ is equal to $\beta S_i(t)/N_i(t)$. Therefore the expected number of newly infected persons in city i at the $t + \delta t$ is:

$$\beta \frac{I_i(t) S_i(t)}{N_i(t)} \tag{3.8}$$

One infected can be recovered during the time period between t and $t + \delta t$ with a probability μ, where mu^{-1} is the typical duration of a decease. Hence, the expected number of newly recovered in city i at $t + \delta t$ is:

$$\mu I_i(t) \tag{3.9}$$

The above formulation assume that the city is isolated and no infected come inside and out the city during the interval $(t, t + \delta t)$. The SIR model can be characterized

by the deterministic set of equations

$$S_i(t+\delta t) = S_i(t) + \Omega(S_i(t)) - \beta\frac{S_i(t)I_i(t)}{N_i(t)} \tag{3.10}$$

$$I_i(t+\delta t) = I_i(t) + \Omega(I_i(t)) + \beta\frac{S_i(t)I_i(t)}{N_i(t)} - \mu I_i(t) \tag{3.11}$$

$$R_i(t+\delta t) = R_i(t) + \Omega(R_i(t)) + \mu I_i(t), \tag{3.12}$$

$$\tag{3.13}$$

where $\Omega(X_i)$ is called *transport operator* and represents the net change in compartment X due to travel from/to city i. From the above system of equations we can extract (by adding them) the conservation equation for the population in each city:

$$N_i(t+\delta t) = N_i(t) + \Omega(S_i(t)) + \Omega(I_i(t)) + \Omega(R_i(t)) \tag{3.14}$$

Using the same formulation above we can write also down the equations that describe the SIS model:

$$S_i(t+\delta t) = S_i(t) + \Omega(S_i(t)) - \beta\frac{S_i(t)I_i(t)}{N_i(t)} + \mu I_i(t) \tag{3.15}$$

$$\tag{3.16}$$

$$I_i(t+\delta t) = I_i(t) + \Omega(I_i(t)) + \beta\frac{S_i(t)I_i(t)}{N_i(t)} - \mu I_i(t) \tag{3.17}$$

$$\tag{3.18}$$

3.4 Numerical Simulations - Results

As we said before we allow the mobility model to run for a period of 1 year in order to avoid early times spreading instability that maybe will caused due to the exploration and preferential return rules.

We choose particular values of the decease parameters β and μ and we initialize the epidemic spreading model by assigning randomly 10 infected agents at the airport of consideration. We first consider the *epidemic threshold* when an epidemic starts from different airports. The epidemic threshold defined by the reproductive number above

which the epidemic causes an outbreak [Fig. 3-5]. It is very clear that the epidemic

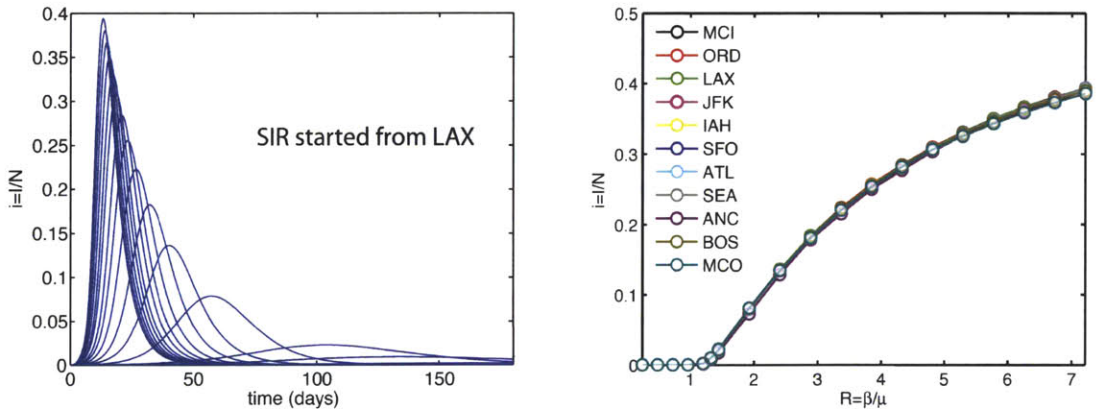

Figure 3-5: SIR Model: *(left)* The density of Infected agents ($i = I/N$) as a function of time for different values of β and $\mu = 0.0412$ when an epidemic started from Kansas City Int. Airport. Each Curve is averaged over 100 realizations and we use 10^5 agents for the mobility model. *(right)* The maximum of the desnsity i_{max} as a function of the reproductive number $R = \beta/\mu$ of the process.

threshold does not depend on the initial position of the infected agents since for all airports the reproductive number above which the epidemic causes an outbreak is the same for all airports and near to $R = 1$ [13].

We next use two quantities in order to quantify early time spreading ability of the different airports: Mean Square Displacement [31] and Entropy [12, 26]:

Mean Square Displacement: This quantity indicates the spatial spreading of the agents at each time and can be defined:

$$\Delta x^2(t) =< (\mathbf{x}(t) - < \mathbf{x}(t) >)^2 >, \tag{3.19}$$

where $\mathbf{x}(t)$ the position vector of all agents at time t and $< . >$ indicates averaging over all the *infected* agents.

Entropy: measure of the level of heterogeneity of decease prevalence and defined

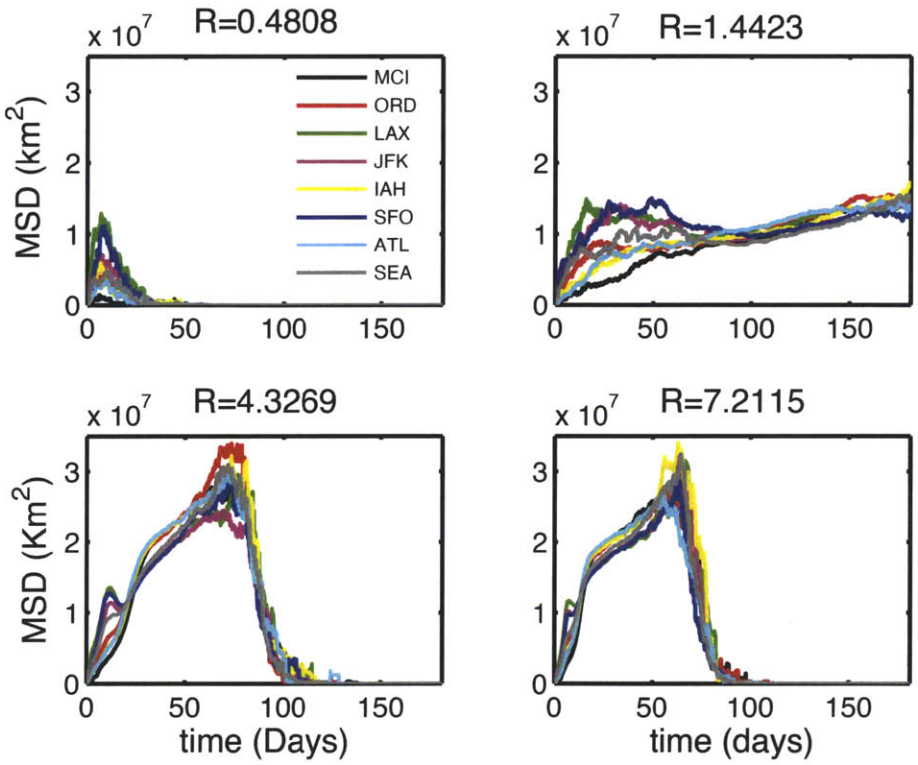

Figure 3-6: SIR Model: The MSD of Infected agents plume as a function of time for different values of $R = \beta/\mu$ where $\mu = 0.0412$ when an epidemic started from different airports in US. Each Curve is averaged over 100 realizations and we use 10^5 agents for the mobility model.

as:

$$H(t) = -\frac{1}{\log V} \sum_j \rho_j(t) \log \rho_j(t), \tag{3.20}$$

where $\rho_j(t) = i_j(t)/\sum_m i_m(t)$ and $i_j(t) = I_j(t)/N_j(t)$ the prevalence in each city j. Entropy takes values between 0 at the initial time when only one city is infected, and 1 when all the airports in the world have the same density of infected agents ($i_m = i_l$ for every m and l).

We plot these two quantities as a function of time in the case of an SIR [Fig. 3-7] and SIS [Fig. 3-8] model for different airports of initialization and try to compare airports in terms of ability to spread in the space the decease. I also report properties like total traffic and connectivity of the airports of consideration [Fig. 3-9].

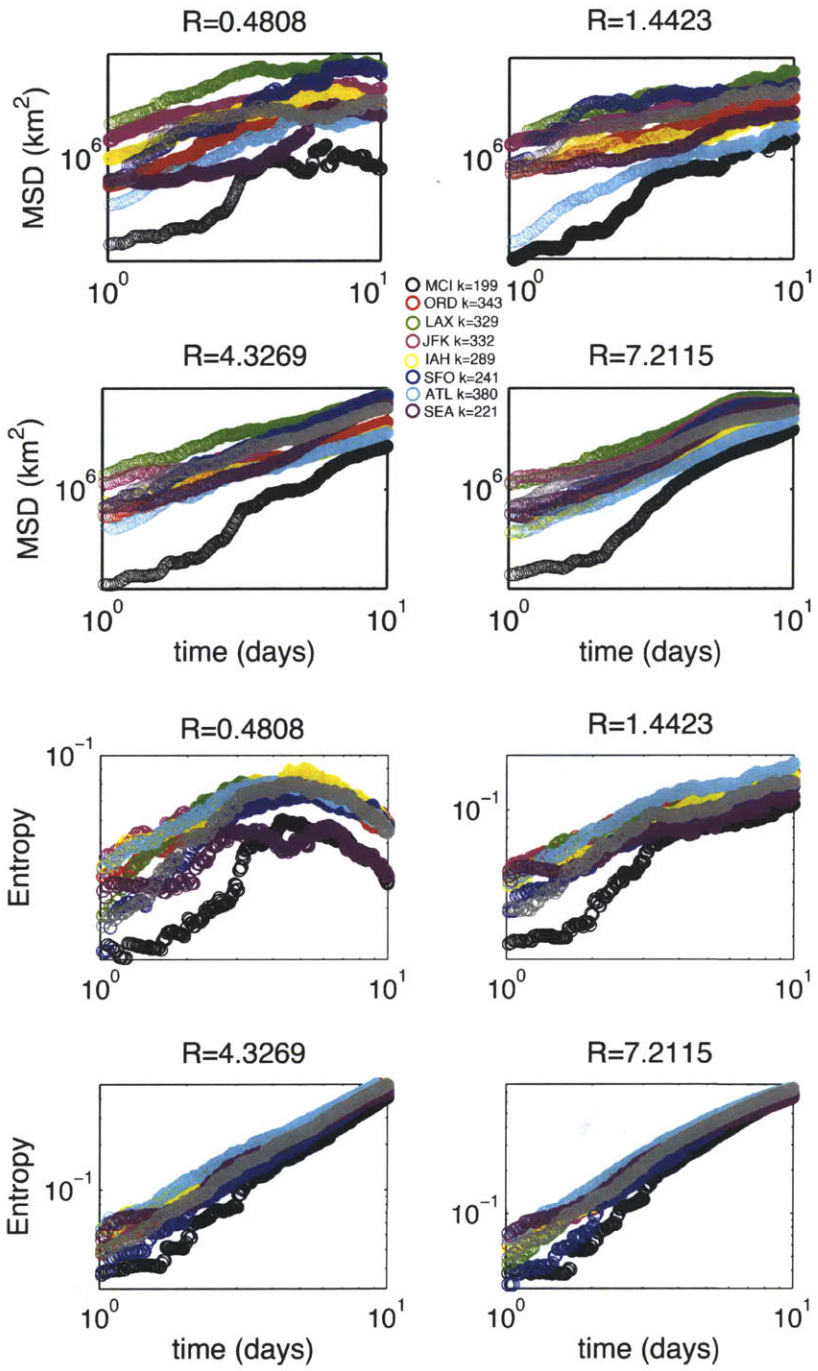

Figure 3-7: SIR Model: The MSD of Infected agent plume as well as the Entropy as a function of time for different values of $R = \beta/\mu$ where $\mu = 0.0412$ when an epidemic started from different airports in US. Each Curve is averaged over 100 realizations and we use 10^5 agents for the mobility model. We concetrate at early time period between 1 and 10 days after the infected appeard in the airport of Consideration

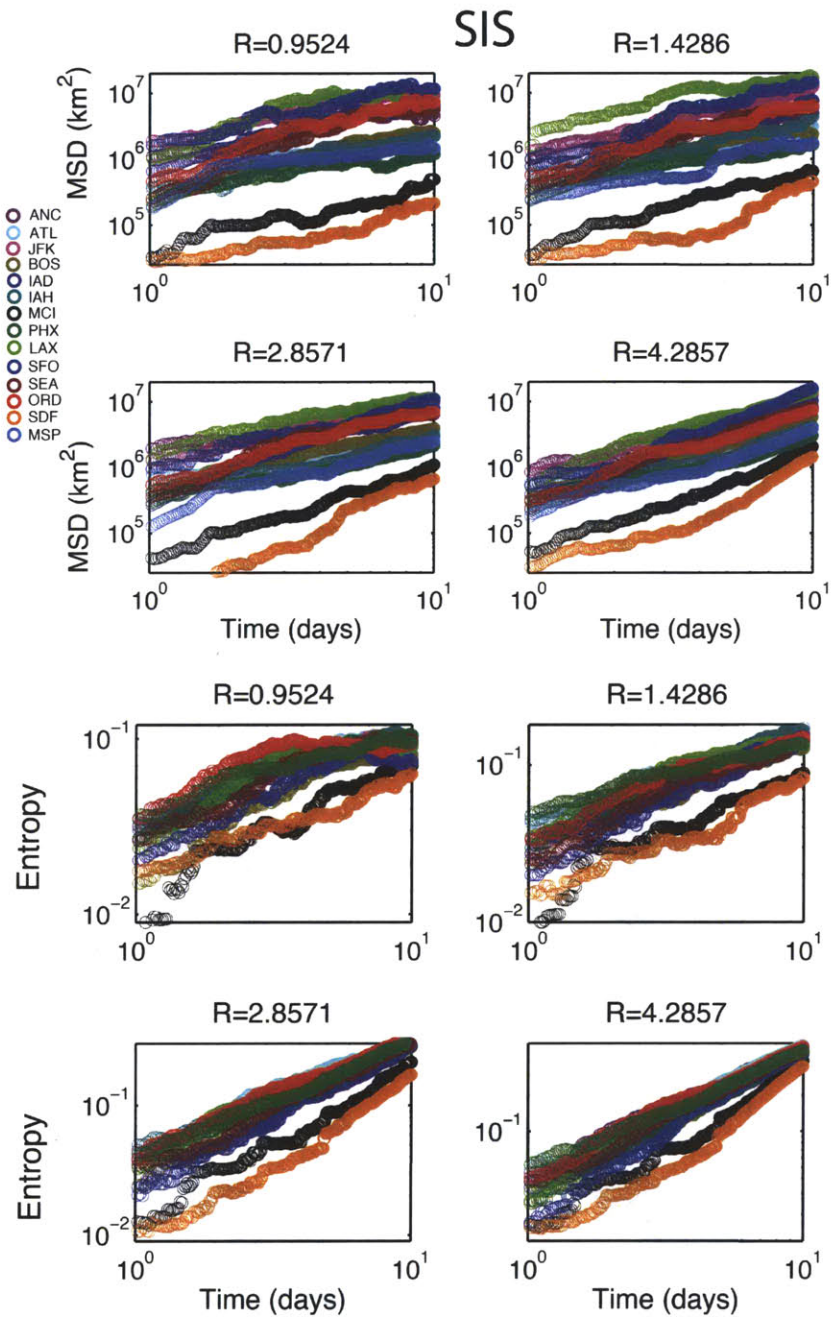

Figure 3-8: SIS Model: The MSD of Infected agent plume as well as the Entropy as a function of time for different values of $R = \beta/\mu$ where $\mu = 0.0412$ when an epidemic started from different airports in US. Each Curve is averaged over 100 realizations and we use 10^5 agents for the mobility model. We concetrate at early time period between 1 and 10 days after the infected appeard in the airport of Consideration

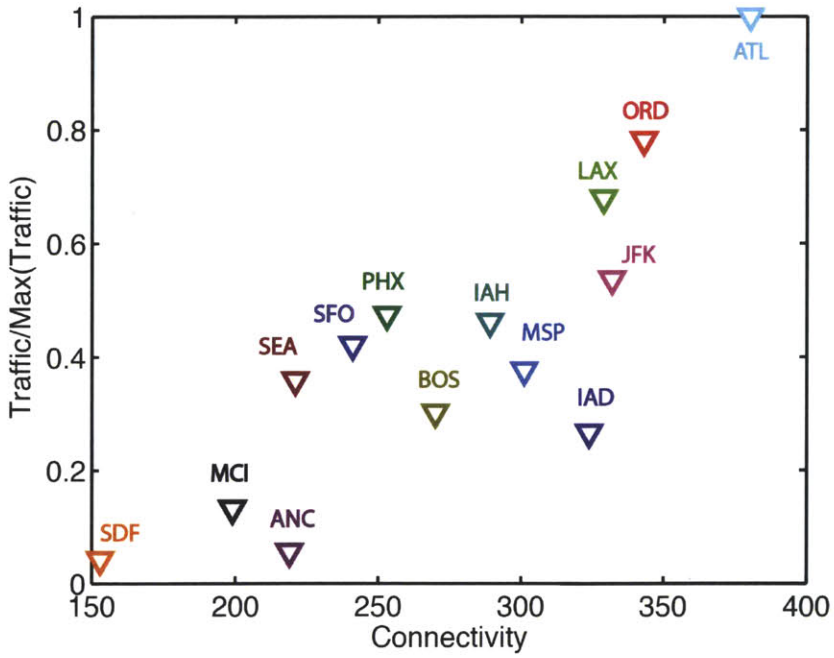

Figure 3-9: Traffic-Connectivity properties of the airports of consideration

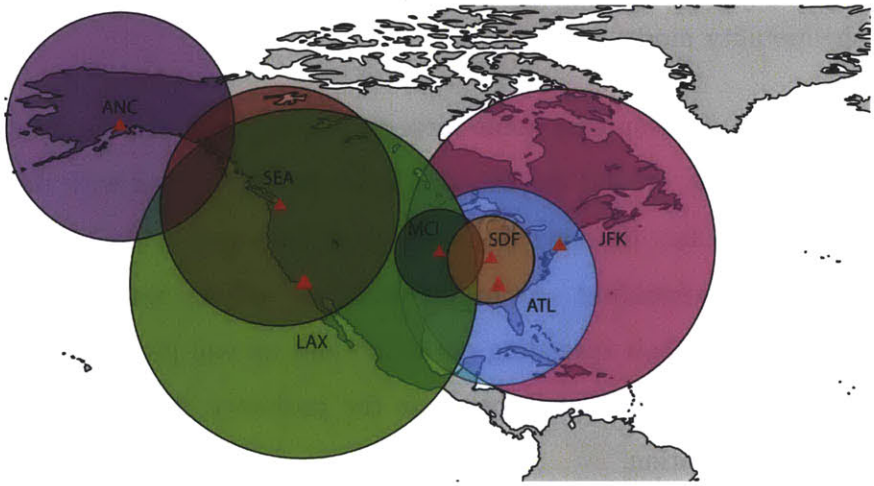

Figure 3-10: The wave front of infected agents 10 days after the epidemic started from Los Angeles International airport. We make use of the SIS model with parameters $\beta = 0.01$ and $\mu = 0.0412$. Results are averaged over 100 realizatiosn and we use 10^5 agents for the mobility model.

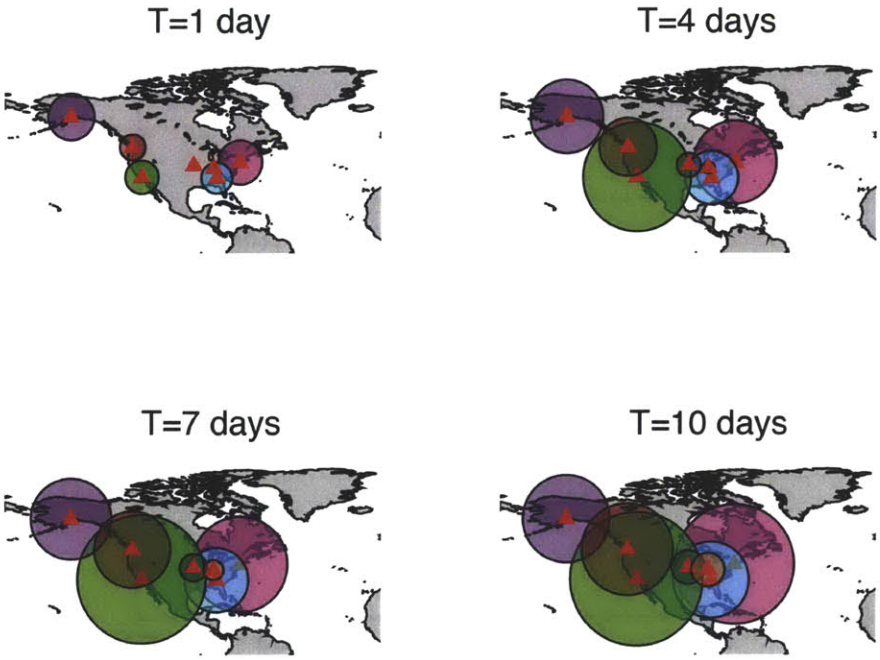

Figure 3-11: The wave front of infected agents at different times after the epidemic started from different airport locations. We make use of the SIS model with parameters $\beta = 0.01$ and $\mu = 0.0412$. Results are averaged over 100 realizatiosn and we use 10^5 agents for the mobility model.

From the above plots it is clear that some airports dominates in terms of ability to spread a decease spatially around the world. This ability correlated with the connectivity and the total traffic. However these two properties cannot define, only by themselves, the ability of spreading. As future work, we will try rank all airports in the database in terms of their spreading potential, and we will propose improved measures of influential spreaders that incorporate the geometry, fluxes through the network and the link orientation.

Chapter 4

Discussion and Conclusions

The nonlinear scaling of spreading with time suggests that transport processes in complex networks are much faster than previously estimated using purely diffusive random walk models that neglect advective fluxes. In the presence of advection, the topology of the network plays a central role in the spreading dynamics, while for diffusion-dominated processes the scaling of spreading with time is independent of network connectivity. Our interpretation in terms of transition times links transport in complex networks with well-established models of effective transport through disordered systems, and opens the door to aggregate conceptualizations of biased transport processes in scale-free networks.

Moreover, the ability of a node to spread an information or a decease in the system cannot be only described by its connectivity or by the total weight of its links (traffic). However, we believe that a combination of properties like connectivity, total traffic, betweenness, k-core degree, spatial orientation and sphericity of the links are essential to define influential spatial spreaders in a complex system at the early times of an outbreak.

The identification of the most influential spreaders, and whether an outbreak will be short-lived or result in a pandemic, depend critically on our ability to predict, in a stochastic sense, the migration and contact dynamics of individuals through the network, which in turn must rely on a faithful characterization of the network topology, geometry, fluxes, and individual mobility patterns.

Bibliography

[1] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proc. of the 32nd Ann. ACM Symp. on Theory of Computing*, Portland, OR, 2000.

[2] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world wide web. *Nature*, 401:130–131, 1999.

[3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[4] C. Barnhart, D. Fearing, and V. Vaze. Modeling passenger travel and delays in the national air transportation system.

[5] A. Baronchelli, M. Catanzano, and R. Pastor-Satorras. Random walks on complex trees. *Phys. Rev. E*, 78:011114, 2008.

[6] B. Berkowitz, A. Cortis, M. Dentz, and H. Scher. Modeling non-Fickian transport in geological formations as a continuous time random walk. *Rev. Geophys.*, 44(2):RG2003, 2006.

[7] B. Berkowitz and H. Scher. Anomalous transport in random fracture networks. *Phys. Rev. Lett.*, 79:4038–4041, 1997.

[8] E. M. Bollt and D. ben Avraham. What is special about diffusion on scale-free nets? *New J. Phys.*, 7:26, 2005.

[9] S. Carmi, Z. Wu, S. Havlin, and H. E. Stanley. Transport in networks with multiple sources and sinks. *EPL*, 84:28005, 2008.

[10] S. Carmi, Z. Wu, E. López, S. Havlin, and H. E. Stanley. Transport between multiple users in complex networks. *Eur. Phys. J. B*, 77:165–174, 2007.

[11] R. Carvalho, L. Buzna, F. Bono, E. Gutiérrez, W. Just, and D. Arrowsmith. Robustness of trans-European gas networks. *Phys. Rev. E*, 80:016106, 2009.

[12] V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. USA*, 68:1893–1921, 2006.

[13] V. Colizza, R. Pastor-Satorras, and A. Vespignani. Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nature*, 2:276–282, 2007.

[14] M. Dentz, H. Scher, D. Holder, and B. Berkowitz. Transport behavior of coupled continuous-time random walks. *Phys. Rev. E*, 78:041110, 2008.

[15] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–60, 1960.

[16] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Comput. Commun. Rev.*, 29:251, 1999.

[17] L. K. Gallos. Random walk and trapping processes on scale-free networks. *Phys. Rev. E*, 70:046116, 2005.

[18] L. K. Gallos, C. Song, S. Havlin, and H. A. Makse. Scaling theory of transport in complex biological networks. *Proc. Natl. Acad. Sci. USA*, 104:7746–7751, 2007.

[19] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[20] P. K. Kang, M. Dentz, and R. Juanes. Predictability of anomalous transport on lattice networks with quenched disorder. *Phys. Rev. E*, 83:030101, 2011.

[21] J. Klafter and R. Silbey. Derivation of the continuous-time random-walk equation. *Phys. Rev. Lett.*, 44:55–58, 1980.

[22] K. C. Leptos, J. S. Guasto, J. P. Gollub, A. I. Pesci, and R. E. Goldstein. Dynamics of enhanced tracer diffusion in suspensions of swimming eukaryotic microorganisms. *Phys. Rev. Lett.*, 103:198103, 2009.

[23] G. Li, L. A. Braunstein, S. V. Buldyrev, S. Havlin, and H. E. Stanley. Transport and percolation theory in weighted networks. *Phys. Rev. E*, 75:045103, 2007.

[24] G. Li, S. D. S. Reis, A. A. Moreira, S. Havlin, H. E. Stanley, and J. S. Andrade, Jr. Towards design principles for optimal transport networks. *Phys. Rev. Lett.*, 104:018701, 2010.

[25] E. López, S. V. Buldyrev, S. Havlin, and H. E. Stanley. Anomalous transport in scale-free networks. *Phys. Rev. Lett.*, 94:248701, 2005.

[26] P. Maynar and E. Trizac. Entropy of continuous mixtures and the measure problem. *Phys. Rev. Lett.*, 106:160603, 2011.

[27] S. Meloni, A. Arenas, and Y. Morenco. Traffic-driven epidemic spreading in finite-size scale-free networks. *Proc. Natl. Acad. Sci. USA*, 106:16897–16902, 2009.

[28] R. Metzler and J. Klafter. The random walks guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339:1–77, 2000.

[29] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms*, 6:161–179, 1995.

[30] E. W. Montroll and G. H. Weiss. Random walks on lattices. II. *J. Math. Phys.*, 6:167, 1965.

[31] C. Nicolaides, L. Cueto-Felgueroso, and R. Juanes. Anomalous physical transport in complex networks. *Phys. Rev. E*, 82:055101, 2010.

[32] J. B. Noh and H. Rieger. Random walks on complex networks. *Phys. Rev. Lett.*, 92:118701, 2004.

[33] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, 2001.

[34] A. F. Rozenfeld, R. Cohen, D. ben-Avraham, and S. Havlin. Scale-free networks on lattices. *Phys. Rev. Lett.*, 89:218701, 2002.

[35] H. Scher and E. W. Montroll. Anomalous transit-time dispersion in amorphous solids. *Phys. Rev. B*, 12(6):2455–2477, 1975.

[36] M. F. Shlesinger. Asymptotic solutions of continuous-time random walks. *J. Stat. Phys.*, 10(5):421–434, 1974.

[37] C. Song, T. Koren, P. Wang, and A.-L. Barabási. Modelling the scaling properties of human mobility. *Nat. Phys.*, 6:818–823, 2010.

[38] M. Stanley. The small world problem. *Phys. Today*, 1:60–67, 1967.

[39] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki. Rules for biologically inspired adaptive network design. *Science*, 327:439–442, 2010.

[40] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small world' networks. *Nature*, 393:440, 1998.

[41] S. R. Yu, M. Burkhardt, M. Nowak, J. Ries, Z. Petrášek, S. Scholpp, P. Schwille, and M. Brand. Fgf8 morphogen gradient forms by a source-sink mechanism with freely diffusing molecules. *Nature*, 461:533–536, 2009.