



## UvA-DARE (Digital Academic Repository)

### Complex networks and agent-based models of HIV epidemic

Zarrabi, N.

**Publication date**  
2013

[Link to publication](#)

**Citation for published version (APA):**

Zarrabi, N. (2013). *Complex networks and agent-based models of HIV epidemic*.

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Introduction

## 1.1 Background and Motivation

We live in a complex world. Many of the world's current problems or natural phenomena can be described as *complex*, from protein-protein interactions [179] and spreading of infectious diseases [56], to social interactions and socio-economics of modern megacities [172, 19], all the way to the human brain itself [159, 31]. The study of complex systems, modeling and simulation provide means for understanding and predicting the behavior of such systems and therefore has the possibility of creating significant impact on understanding the complex world around us.

A complex behavior can occur in any system that consists of large numbers of components which interact non-linearly [11]. A complex system can also interact with its environment and have emergent properties, such as molecular and cellular systems, organisms, ecosystems and human societies. To study a complex system, it is essential to start with real-world data as a base for building models. A model is an abstraction of a real system which is based on a limited number of assumptions about the “physical” behavior of the system. Models can be used for understanding the connections between the system's state variables (input, internal and output variables) and assessing the system's behavior. If fed with the right data, modeling approaches can

also provide the requested level of predictability in many complex systems [172].

In the field of biology and bio-medical sciences, many models are developed to help a better understanding of the underlying biological phenomena and medical processes. Let us consider the problem of infectious disease spreading in human societies. An infectious disease results from the presence and increasing the number of pathogenic biological agents (i.e. viruses, bacteria and microbial pathogenic agents) in a host organism. Social interactions as well as genetic diversity of the transmitted viral agent among individuals dictate the dynamics of infectious disease spreading in a population. Hence, the infection transmission can be investigated at different spatio-temporal scales, from molecular to epidemiological levels [150].

Human immunodeficiency virus (HIV) infection is one particular example of an infectious disease that can be mapped to a multi-scale system. The genetic evolution and mutations of the virus occurs at a molecular/cellular scale within a time span of milliseconds. On the other hand, the transmission and spread of the virus is very slow and happen mainly through blood exchange or sexual intercourse between individuals in a population. Many data has been collected from HIV infected individuals, from genome sequencing and blood samples to contact tracing and social or sexual interactions. The data from each scale provides information on the dynamics of the system on micro- or macro-level system dynamics. Models that are built only based on macro-scale data, neglecting the micro-level dynamics, might be too abstract and not be accurate enough in representing the real system. For models that are built only based on micro-scale data, there might be lack of understanding of the underlying processes of the system and such models are most of the times not efficient.

Complex systems around us are multi-scale and are made up of interacting agents with heterogeneous behaviors and goals. Modeling these systems poses great challenges for understanding and predicting their dynamics. New modeling approaches are necessary to handle the multi-scale nature of such complex systems. To model, quantify and analyze HIV dynamics, there is an urgent need for methods that can capture the multi-scale spatio-temporal characteristics of complex

systems. Data-driven modelling is based on the analysis of all the data characterising the system under study. Formulating data-driven models that couple together data present at different scales helps us to address the multi-scale properties of a complex systems. In what follows, we explain the HIV epidemic and the urge for computational and modeling methods to tackle the epidemic.

### 1.1.1 HIV Epidemic

HIV is a member of the retrovirus family which causes acquired immunodeficiency syndrome (AIDS). HIV reduces the number of immune cells in an infected body and causes the immune systems to lack the ability to fight off other infectious diseases. The first symptoms of the virus was first observed in 1980s and since then HIV infection and AIDS have turned out to be one of the most destructive epidemics of mankind. AIDS has so far killed more than 25 million infected people around the world [1, 73] and globally, at least 34 million people were living with HIV at the end of 2011 [2].

Until now, there is still no exact cure for AIDS but there are anti-retroviral drugs developed to control the progression of HIV and to make the lifespan of patients longer. There are different classes of antiretroviral agents that act at different stages of HIV replication cycle. A combination of these drugs is used, known as highly active anti-retroviral therapy (HAART), has had a major impact on suppressing HIV viral replication and transmission. However, HAART has severe side effects and increases the chance of acquiring drug-resistant viral strains [63].

Some virological or medical experiments may be expensive or time-consuming to conduct, which makes them infeasible to study. For example progression of HIV infection is slow and development of AIDS is a long term process (approximately 10-12 years under antiretroviral therapies) [125]. In this case, conducting medical experiments to study the dynamics of HIV is time-consuming and infeasible. Therefore, computer models of HIV dynamics and immune system response could be a time-efficient and cost-effective approach to study this problem. These models assist researchers and pharmacologists to study the

viral dynamics and investigate the effect of different drugs on disease progression (ViroLab, the virtual laboratory for decision support in viral disease treatment [148]).

### 1.1.2 HIV Dynamics

The dynamics of HIV occur from intracellular viral replication within a single cell, to mutations of viral strains within a single individual, to transmission of the infection between individuals. These dynamics occur at different spatio-temporal scales (i.e. molecular/cellular, individual and population) and therefore have different timing and spatial properties. However, the dynamics at all scales are mutually coupled and drive in large-scale dynamics of HIV epidemic as a whole.

The dynamics at molecular and cellular scales involve many different molecular interactions. A single cell itself has a complex heterogeneous environment. The cell consists of many interacting proteins and molecules which arise complex behaviors such as replication, internalization, and movement using motor-molecules. With respect to HIV infection, these dynamical processes become more complicated to study. The interaction of virus-host cells (mainly T lymphocytes) and intracellular dynamics of HIV such as intracellular transport of molecules [51, 27], viral kinetics [164, 134] and intracellular replication of the virus [137, 185] are examples of such processes.

At individual scale, viral dynamics generates  $10^{10}$  virions every day, coupled with a high mutation rate of approximately  $3 \times 10^{-3}$  mutations per nucleotide base cycle of replication and recombinogenic properties of reverse transcriptase [187]. This scenario leads to the production of many variants of HIV in a single infected patient in the course of a day [45].

At population scale, we live in an interconnected world with advanced transportation and mobility infrastructure [172, 173, 19]. Thus, an infected individual can travel and carry the infection from one part of the world to the other part over the course of a single day [56].

## 1.2 Modeling HIV Dynamics

Due to multi-scale dynamics of HIV, modeling and simulation of HIV dynamics can vary from modeling a single cell infected by a virus to modeling the entire immune response (under virus infection and drug therapy) [101] up to simulating the transmission of the virus within a population in a society [152, 104]. Most of the initial modeling research contributed to the study of transmission dynamics of HIV at the population level [125, 111], since little was known about the epidemiology of the virus. Later, with development of genome sequencing technology an enormous amount of data has been gathered on the genetic sequences of viral strains. Molecular scientist use bottom-up approaches such as phylogenetic tree analysis to infer the transmission and evolution of HIV using these data.

In recent years, many mathematical and computational models have been developed to investigate the complexity of HIV dynamics, immune response and drug therapy [182, 111, 161, 35, 69, 149]. For example at the molecular and cellular scales, the dynamic aspects of a cell such as gene transcription, translation, and protein-protein interaction has been modeled using numerical and computational techniques [137, 185]. Such models become more complicated when the cell is infected with a virus and is going under viral replication. At the epidemiological scale, scientists have been trying to study the spread of infectious diseases using social or sexual contact networks, modelling the population as a complex network (where nodes are individuals and links are relationships) and running models of disease spread on top of that. In the case of type HIV-1 infection, these models have been used to understand the complexity of HIV-1 transmission and spread of viral drug resistance [52, 176, 158, 157, 104, 171, 152]. However, many simplifying assumptions are being made on these models as well as on the network structure and topology. For instance, a basic SIR model which classifies individuals into three distinct groups, Susceptible (S), Infectious (I) and Recovered (R), have been used to model the spread of disease. Even though such a mathematical model provides a general framework to understand the transmission of disease in a social network, it may be too simple to accurately model a real

epidemic [171].

Current modeling approaches mainly focus on each scale separately by taking the data from either molecular, individual or population scales. The main modeling methods that are used are either bottom-up or top-down approaches. Phylogenetic analysis, agent-based modeling and complex networks are perhaps the most prominent examples of such methods. Phylogenetic analysis is based on the genetic sequence data. Agent-based models (ABMs) are built based on individual-level behaviors (also known as micro-level dynamics), while complex networks provide global-level properties (or macro-level dynamics) of the system. In what follows, we briefly introduce methods that are being used for studying HIV dynamics.

### 1.2.1 Phylogenetic Analysis

Phylogenetic theory exploits genetic information of viruses using mathematical methods of evolution [92, 58, 90]. Phylogenetic trees are binary hierarchical trees that show the evolutionary relationships among genetic sequences in a population, where topology and branch lengths are estimated via likelihood-based or distance-based methods. Genetic sequences are placed at the leaves of these trees and the internal nodes are considered as hypothetical ancestors under a species' coalescence paradigm. Phylogenetic trees can be used to infer transmission clusters, temporal and geographical dynamics. However, these methods may not necessarily accurately represent the evolution of species, both due to strong assumptions of the underlying mathematical models, and due to noise in the data. For instance, evolution of species is not always reducible to a tree form and a hierarchical tree may not necessarily accurately represent the evolution and transmission of disease. Nonetheless, phylogeneticists have been using extensively phylogenetic analysis to investigate transmission events of infection [104], especially for viruses such as HIV that have a high genetic diversity. However, the agreement between phylogenetic reconstruction and epidemiological evidence of transmission events can be decreased due to other species-specific factors: in the case of HIV infection, those include the long period of infectivity and sparse time and space sam-

pling [138, 79, 28].

### 1.2.2 Agent-based Models

Agent-based models (ABMs) consist of large numbers of heterogeneous entities, known as agents, that interact with each other according to some rules; through the interaction of the agents, system level phenomena are said to emerge. In the past 20 years ABMs have been used to investigate a number of different complex systems, from traffic and parking within cities [13, 18, 55] to cellular interactions and immune system dynamics [22, 185, 101]. ABMs are especially useful for simulating the dynamics of those systems that are driven by human behavior, such as social systems [44], financial markets [23], economics [74] and pandemics [56].

Depending on the system under consideration the agent can be defined through a simple set of rules, or a more sophisticated entity with many interacting rules governing its behavior. In a complex social system, the definition of autonomous decision-making agents can be an abstraction of human actions in the system. ABMs are often built by first specifying the system components, compiling relevant information about entities at a lower level of the system and formulating theories about their behavior. The theories are then implemented in a computer simulation and the emergence of system-level properties can be observed [67]. Agent-based modeling provides a means to incorporate individual-level dynamics in studying complex systems.

### 1.2.3 Complex Network Models

Complex Networks have provided insight for understanding many complex phenomena [162, 177, 152, 157]. A complex network forms when the components of a system are linked and have dynamical interactions. Many real-world applications can be described as a complex network, such as social networks [116], the network of protein-protein interactions [179] and the World Wide Web [6]. These networks are huge with some having thousands or millions of nodes. Besides their enormous size, what make these networks complex is the dynamics of

interactions, which lead to a particular arrangement or topology of the network elements.

A complex network can be described as a graph composed of a set of vertices and a set of edges. The number of edges that connect from a node to other nodes in the network is called the *degree* of the node. The frequency distribution of degrees over the whole network is an important characteristic of the network, called *degree distribution*. The connectivity of all the nodes in the network is characterized by the degree distribution. There are various different forms of complex networks, including scale-free [15], small-world [178] and random networks [113]. A scale-free network is a form of network with a degree distribution that follows a *power law* of the form  $P(k) \sim k^{-\gamma}$ , where  $k$  is the degree,  $P(k)$  is the probability of a node with degree  $k$ , and  $\gamma$  is the *exponent* of the power law. A power-law distribution implies that the majority of nodes in the network have a low-degree, but also that there are a few high-degree nodes, known as *hubs* [114, 40]. Power law distributions and scale-free networks occur in a wide range of phenomena [15, 145]. Small-world networks are another form of network in which most pairs of vertices are connected through a short path. The small-world effect is not confined to social networks and is also observed in many other networks such as brain networks [31] and the electric power grid networks [3]. Random graphs are the simplest form of complex networks, in which every pair of nodes are connected randomly with an independent probability  $p$ . The degree distribution of such a graph is then a Poisson distribution. Random graphs however, typically do not produce the topological and structural properties of real-world systems observed in nature and society [5, 20].

Much of the research in the area of complex networks focuses on examining real-world systems and understanding what form of network structure the particular natural phenomena maps to. The scale-free structure and power-law distributions are known to occur in many real-world networks. An example is the network structure of sexual contacts of homosexual males in a population, which is known to be scale-free with a power law exponent value ranging between 1.5 to 2.0 [145, 104]. The high-degree nodes in the sexual networks are the promiscuous individuals who may accelerate the spread of

sexual transmittable infections in a population. Another example is the network of protein-protein interactions, which is also known to have a scale-free structure [83]. Highly-connected nodes in such networks correspond to important proteins that play a major role in a biological functions. To predict the behavior of a complex system, it is essential to start with the mathematical description of patterns observed in real-world data [172]. The definition of complex networks and knowledge of network characteristics provide such descriptions for many real-world systems. The degree distribution is an example of a mathematical formulation for the connectivity patterns in the network. Knowledge of network characteristics (i.e. degree distribution, average path length, centrality, clustering coefficient and community structure) provides insights on interconnectivity of components and global-level properties of a system [14, 20].

### 1.3 Thesis Overview

In this thesis, we explore the limits of these multi-scale models by looking into the HIV data present at different scales (from molecular and cellular to epidemiological scales). We build data-driven models and perform network analysis in order to understand the dynamics of HIV epidemic at different spatio-temporal scales. We also introduce and define a new modeling methodology called “Complex Agent Networks” (CANs), by combining complex networks and agent-based modeling techniques. The introduction and use of CANs helps to better capture and depict the characteristics of a complex system at both individual and global system levels. In addition we think that the multi-scale nature of CANs offer a natural way of absorbing complex, multi-scale real world data. These systems have sparked major attention and facilitated substantial applications in scientific fields, e.g., ecosystems, epidemiology, financial markets and economics.

In chapter 2, we propose a computational model of HIV intracellular replication where infected cells undergo a single cycle of virus replication. A cell is modeled as an individual entity with certain states and properties. The model is stochastic and keeps track of the main viral proteins and genetic materials inside the cell. Two simu-

lation approaches are used for implementing the model: rate-based and diffusion-based approaches. The results of the simulation are discussed based on the number of integrated viral cDNA and the number of viral mRNA transcribed after a single round of replication. The model is validated by comparing simulation results with available experimental data. Simulation results give insights about the details of HIV replication dynamics inside the cell at the protein level. Therefore the model can be used for future studies of HIV intracellular replication in vivo and drug treatment.

In chapter 3, a novel method is introduced to reconstruct HIV transmission networks based on patients genetic, demographic and clinical data. The method is based on real patient data and considers epidemiological factors as well as viral genome data for network construction. We argue that combining data from different scales is required for a more realistic description of complex systems behavior such as transmission of infectious disease and HIV epidemic.

In chapter 4, we modify the filter-reduction method presented in chapter 3 to study the transmission of HIV in mixed risk groups.

In chapter 5, we introduce CANS, as a new modeling methodology for modeling complex systems. We argue that CANS are able to capture both individual-level dynamics as well as global-level properties of a complex system, and as such may help to obtain a better understanding of the fundamentals of such systems.

Chapter 6 concludes the thesis and presents future perspectives of this research.