



Information and Communication Technologies

EPIWORK

Developing the Framework for an Epidemic Forecast Infrastructure

<http://www.epiwork.eu>

Project no. 231807

D2.2 Theoretical foundation and mathematical description of network- network systems

Period covered: 13th – 24th months
Start date of project: February 1st, 2009
Due date of deliverable: February 20th,
2010
Distribution: public

Date of preparation:
Duration:
Actual submission date:
February 20th 2010
Status: Final

2 | EPIWORK D2.2

Project Coordinator: Alessandro Vespignani

Project Coordinator Organization Name: ISI Foundation

Lead contractor for this deliverable: Max Planck Gesellschaft zur Foerderung der Wissenschaften E.V.

Work package participants

The following partners have taken active part in the work leading to the elaboration of this document, even if they might not have directly contributed writing parts of this document:

- MPG
- ISI
- FGC-IGC
- TAU
- BIU
- FBK

Change log

Version	Date	Amended by	Changes
1	23/02/11		

1 Theoretical Foundation

This deliverable is directly linked to the successful completion of Tasks 5 and 6 defined in Theme II of WP2. The key goal is to bridge the conceptual gap between individual based network models in which nodes reflect individuals and their connections represent potential transmission pathways and meta-population models in which each node reflects entire populations and the exchange of individuals their connections. A key problem is that although data exists on both levels, for example passenger or commuter traffic between populations as well as contact patterns between individuals in single populations, comprehensive data on both aspects is unavailable. However, instead of comprehensive inter-individual activity across a meta-population, variability in individual activity is more accessible to measurement.

We therefore developed a theoretical framework that is based on individual activity patterns. This framework is based on the notion that a population can be grouped into subgroups of individuals that share common activity patterns. A group labeled i is characterized by activity a_i , the rate at which individuals have contacts with other individuals. The framework is set up such that these subpopulations represent compartments of single populations or multi-level subsets of hierarchies of populations.

Since disease dynamics requires transmission and therefore interactions between individuals a key task is to infer interaction profiles from measureable activity profiles of individuals or groups of individuals. In order to better understand the dynamic consequences of different ways of modeling the relation between activity and interactions we devised a mathematical approach in which two qualitatively different scenarios can be accounted for. Each scenario reflects extreme cases of a linear spectrum of possibilities.

1.1 General Modelling approach

The foundation of the model is based on the assumption that the interaction rate between individuals of two groups only depends on the activity patterns of each group, this dependence however can take different functional forms. A detailed description of the modeling approach is provided in the supplementary document. The model is derived from a general

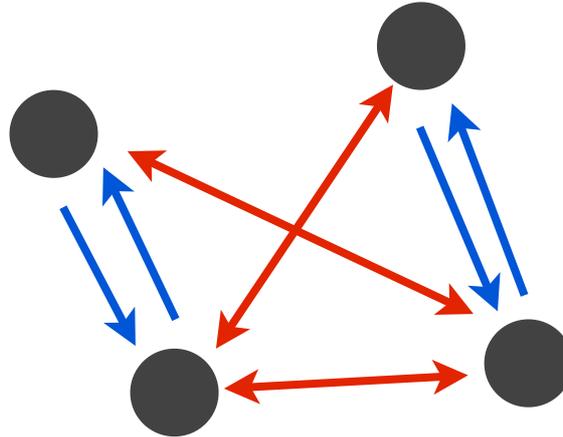


Figure 1: Disease dynamics in single populations with variable activity patterns. *Top left:* Individual activity can be related to interaction rates in two qualitatively different ways, multiplicative (red) and additive (blue) interactions.

metapopulation model with two or more compartments that reflect the infectious state of individuals, e.g. susceptible S and infectious I . Additionally individuals have a discrete or continuous index that distinguishes groups of individuals according to their activity patterns. Formally the dynamics of e.g. SIS or SIR models are then determined by a matrix k_{nm} that quantifies the number of interactions between two groups of labeled n and m . The date of the dynamical system (e.g. the stability of the disease free state) is governed by spectral properties of the interaction matrix k_{nm} . Although this has been known for some time, it has remained unclear how to map activity patterns of groups of individuals onto the interaction patterns between groups. The model we developed indicates that the stability of the disease free state, the dynamics of the entire systems, the magnitude of fluctuations and absorption probabilities depend strongly on the way the model captures how individual activity patterns are mapped onto interactions. This is illustrated by the two extreme cases discussed below and explained in detail in the supplementary document.

1.1.1 Multiplicative Interactions:

In this scenario, the rate of interaction of an individual with activity a_i and another individual of activity a_j is essentially given by the product:

$$k_{ij} = \frac{2a_i a_j}{\langle a \rangle},$$

where the denominator reflects the mean activity in the entire meta-population. This functional form corresponds to situations in which a contact of individuals is negotiated and mutually anticipated. For example interactions triggered by dates, meetings and non-anonymous activities. Contemporary network models implicitly model interactions by multiplicative coupling. In heterogeneous populations this approach leads to a stability criterion that depends on the mean basic reproduction ratio and its coefficient of variation:

$$\langle R_0 \rangle (CV_{R_0}^2 + 1) \geq 1.$$

This relation is equivalent to a similar criterion in static network models in which the degree of a node is related to the number of possible infection pathways:

$$\langle k^2 \rangle / \langle k \rangle \geq 1.$$

Both relations are the theoretical foundation of the absence of epidemic thresholds in scale free networks or populations with strongly heterogeneous activity patterns and that epidemic thresholds decrease with variability in activity. However, the important CV^2 rule above only holds for this specific assumption and not in general.

1.1.2 Additive interactions:

In fact, many disease dynamical interactions between individuals are more appropriately described by additive interactions. In this scenario the interaction rate is given by the sum of activities of each individual:

$$k_{ij} = a_i + a_j$$

These types of contacts reflect situations that are encountered in more anonymous situations, for instance contacts in public transportation systems, crowds, concerts and airport gates. The theoretical framework that we

developed extends this approach to incorporate additive interactions as described above that yield a qualitatively different relation for the threshold given by:

$$\frac{1}{2}\langle R_0 \rangle \left(1 + \sqrt{CV_{R_0}^2 + 1}\right) \geq 1$$

This relation is substantially different from the known relation above and

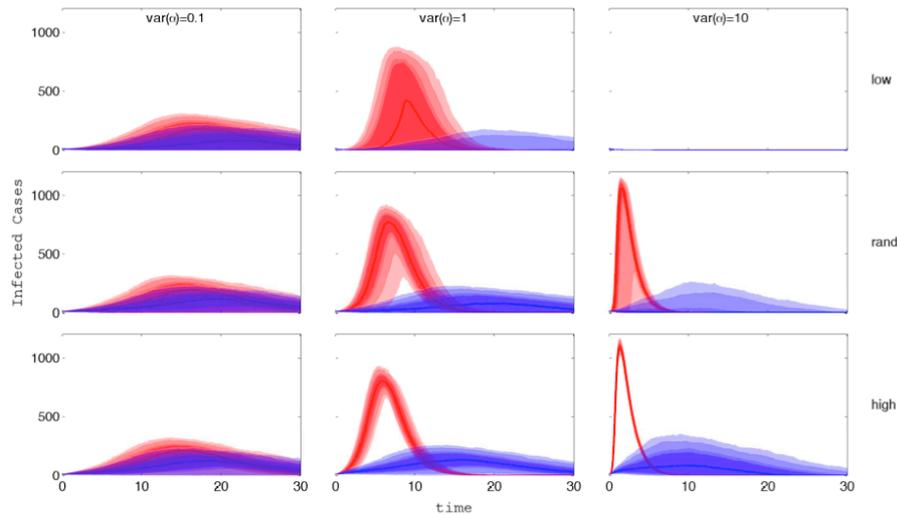


Figure 2: If a population is characterized by a distribution $p(a)$ of activities, different models for mapping activity to interactions can yield strong differences in the epidemic time course as indicated by the trajectories of stochastically simulated SIR models with additive (blue) and multiplicative (red) interaction rates.

implies a very critical behavior in parameter space (see Fig 3). Although this relation implies the decrease of the epidemic threshold with variability in activity patterns the functional dependence is substantially different and indicates that epidemic thresholds can be over- or underestimated if only multiplicative interactions are assumed. Only in homogeneous populations both types of mappings coincide. The difference in epidemic thresholds in additive and multiplicative scenarios is depicted in Fig. 2 and 3. In the case of a stochastic SIR model, the typical time course of epidemics can vary over orders of magnitude, as is illustrated in Fig. 2 which compares various realization of SIR trajectories driven by multiplicative or additive coupling.

1.2 Network Network Duality Model

Based on both modeling paradigms we devised the mathematical foundation in which both interaction scenarios can be accounted for beyond the meta-population ansatz. We obtained a systematic mapping from possible interaction pathways between individuals n and m are defined by an adjacency matrix A_{nm} the dimensionality of which is given by the size of the population. Additionally groups of individuals are categorized according to their activity profiles. The interaction between different groups can be

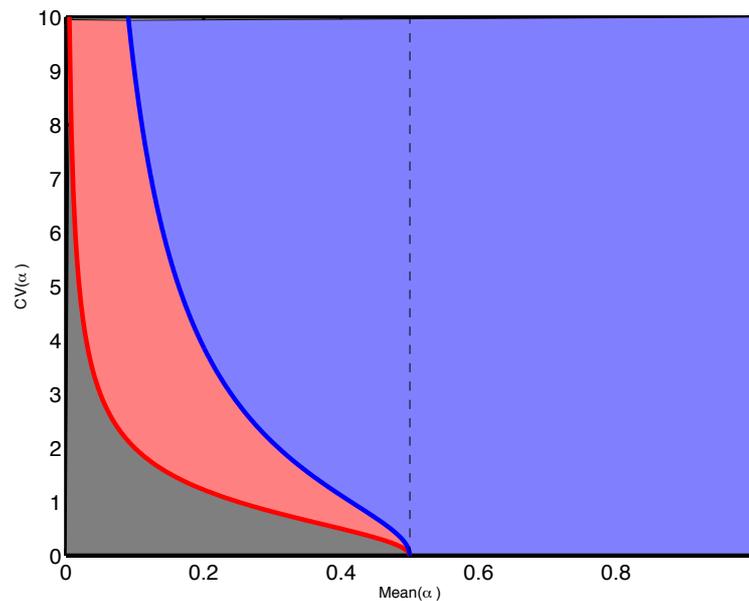


Figure 3: The dependence of the epidemic threshold on mean basic reproduction ratio and coefficient of variation is depicted on the right. In the grey region the system is below the critical threshold. Above the red and blue line multiplicative and additive interaction systems cross the threshold, respectively. Not that even basic reproduction numbers well below the critical threshold in the homogeneous systems ($\alpha=0.5$), strong variability in contact rates destabilizes the system but much earlier in the multiplicative model.

computed as a function of the inter-individual connectivity given by the adjacency matrix, the vector of activity profiles of subpopulations and the nature of interactions (additive vs. multiplicative) between subgroups. Effectively, the dynamical system is projected onto a lower-dimensional dynamical systems in which the dynamical variables are the fraction of infected individuals in a subgroup similar to conventional meta-population models. However, the approach we developed provides a road map for integrating adjacency on the individual level as well as variability in activity profiles among groups of individuals. Details of this model are provided in the

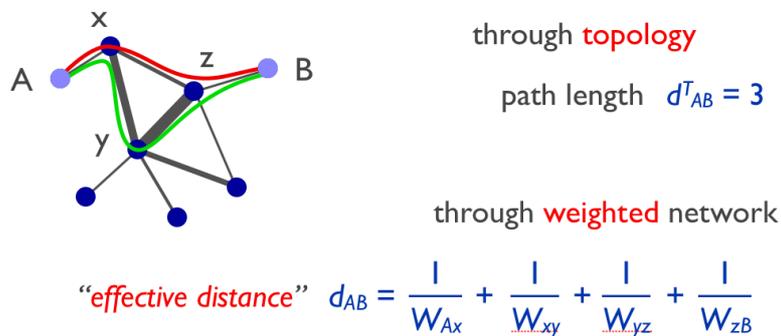
supplementary document (a poster presented at the Netsci 2009). We have tested the modeling framework in sets of artificial settings in order to investigate who dynamical processes depend on a) the statistical properties of inter-individual adjacency b) the distribution of activity profiles and c) the nature of interactions. However we have not yet gauged the data on respiratory diseases in the UK as anticipated. This is due to the change in focus of the entire WP2 during the active modeling phase of the H1N1 pandemic in 2009, which required full attention of the participating partners in WP2.

2. SPATO - Shortest Path Tree Tomography

On the interface of deliverable D2.2 with Themes 1 and 3 of WP2 as well as across work packages with WP4 we developed an interactive network analysis and visualization tool. This software tool has been designed to visualize and investigate mobility and transportation networks that resemble the connectivity structure among populations in large scale meta-populations and is therefore of relevance to deliverable D2.2. In fact it is anticipated that based on the network-network model described above social contact networks can be analyzed and visualized with the software as well as dynamical processes that evolve on these networks. This interactive software named SPATO was developed based on the a computational technique developed by us that is based on shortest path tree tomography. The fundamentals of this computational technique are illustrated in Fig. 4.

Both network types that are involved in transportation and mobility on one hand and in social contact networks on the other, are typically strongly heterogeneous. In mobility networks this heterogeneity is reflected in the traffic flux elements W that represent the flux of traffic between populations and that typically vary over many orders of magnitude. In social networks the interaction frequency of individuals n and m represent the connectivity strength and vary over many orders of magnitude as well.

Shortest Paths



Shortest-Path-Tree Tomogram

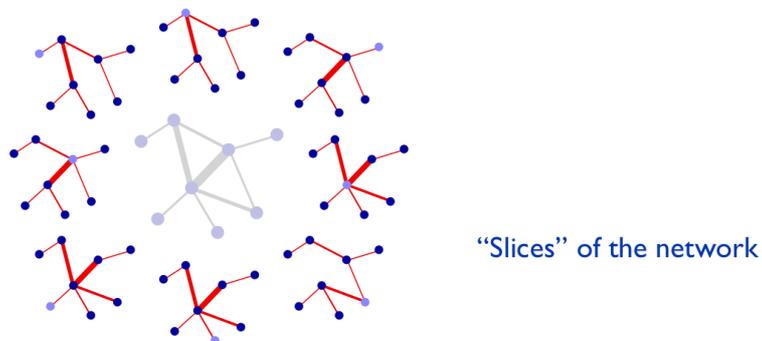


Figure 4: Shortest path based on weight reciprocity. Top: The schematic illustrates the different between shortest “topological” paths and effective paths based on the inverse coupling strength w . The effective distance between nodes A and B is given by the the sum of the inverse coupling strengths along the shortest path. Bottom: For each root node (light blue) in a network (center) the collection of shortest paths to all other nodes defines a shortest path tree for each node, the basis for SPATO visualizations.

Shortest path tomography is based on two essential properties of networks or this nature. It is based on the notion of the perspective of a given node onto the rest of the network and that a weight connecting two nodes is inversely proportional to an effective distance that reflects a meaningful metric in such networks. The notion of effective distance implies that if two nodes are strongly coupled (high traffic in transportation networks and high contact rates in social networks) their effective distance is small, if on the other hand two nodes are only weakly coupled their effective distance is large. Given a chosen root node, a shortest path to another node is defined as the set of links in the network that minimizes the total effective distance given by the

segments involved in the path. A shortest path tree rooted at a given node is provided by the collection of shortest paths to all the other nodes. Therefore, each node has a unique shortest path tree. SPATO allows the user to choose a given root node and displays its shortest path tree, its “perspective” on the network and shows the effective distance to all the other nodes in various two dimensional representations, see. Fig. 5.

SPATO reads origin-destination data that define mobility networks,

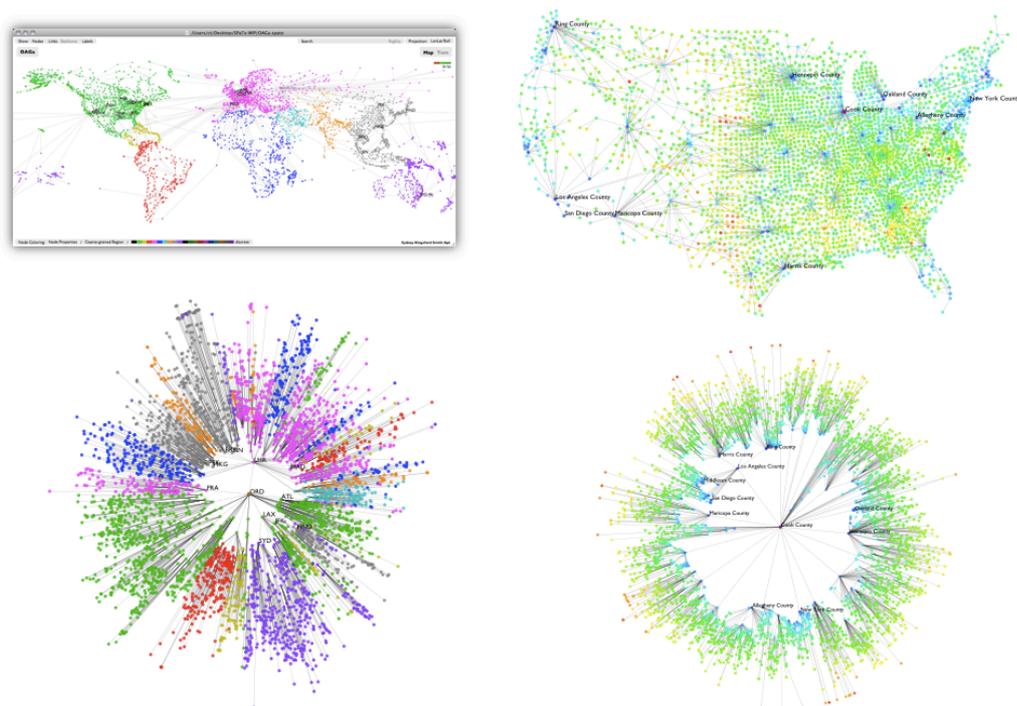


Figure 5: Multi-scale mobility networks as visualized by shortest path tree tomography. Each panel depicts screenshots of the interactive network visualization tool SPATO that was developed in WP in order to visualize disease dynamical processes on a global scale, extract generic features of the dynamic process and relate them to topological features of the underlying mobility network

automatically computes network characterizations, modularity structures, centrality measures of interest and statistical features of a given network. The visualization algorithm of SPATO extracts mobility networks from their geographical embedding and visualizes the network based on shortest paths and effective distances. The interactive tool permits the visualization of

network characteristics, node specific parameters and allows to investigate the perspective of mobility networks from a node of the user's choice. SPATO is currently available as a testable beta version, was implemented in the open source JAVA class Processing and will be made available to the partners of the program once a usable data interface has been implemented. Fig. 5. illustrates screenshots of SPATO for various networks. SPATO has been designed to be usable by a broad spectrum of users, understands widely accepted formats (XML, Excel, CVS) to read in origin destination data and data on dynamical processes that evolve in these networks. In meta-population models SPATO offers a novel way to show global disease dynamics which so far have only been visualized in the contemporary geographical representation. Analyses on toy simulations show that SPATO represents a much more meaningful representation, as discussed below.

2.1 Effective Dynamics on complex multi-scale mobility networks

One of the key motivations in the development of shortest path tomography (SPATO) was the need to understand the complexity exhibited in spatio-temporal patterns of disease dynamics. In contemporary visualizations, disease dynamics are typically displayed geographically, i.e. a geographical projection. In these geographic representations dynamic patterns typically exhibit non-local and multi-scale fractal structure. However, contemporary

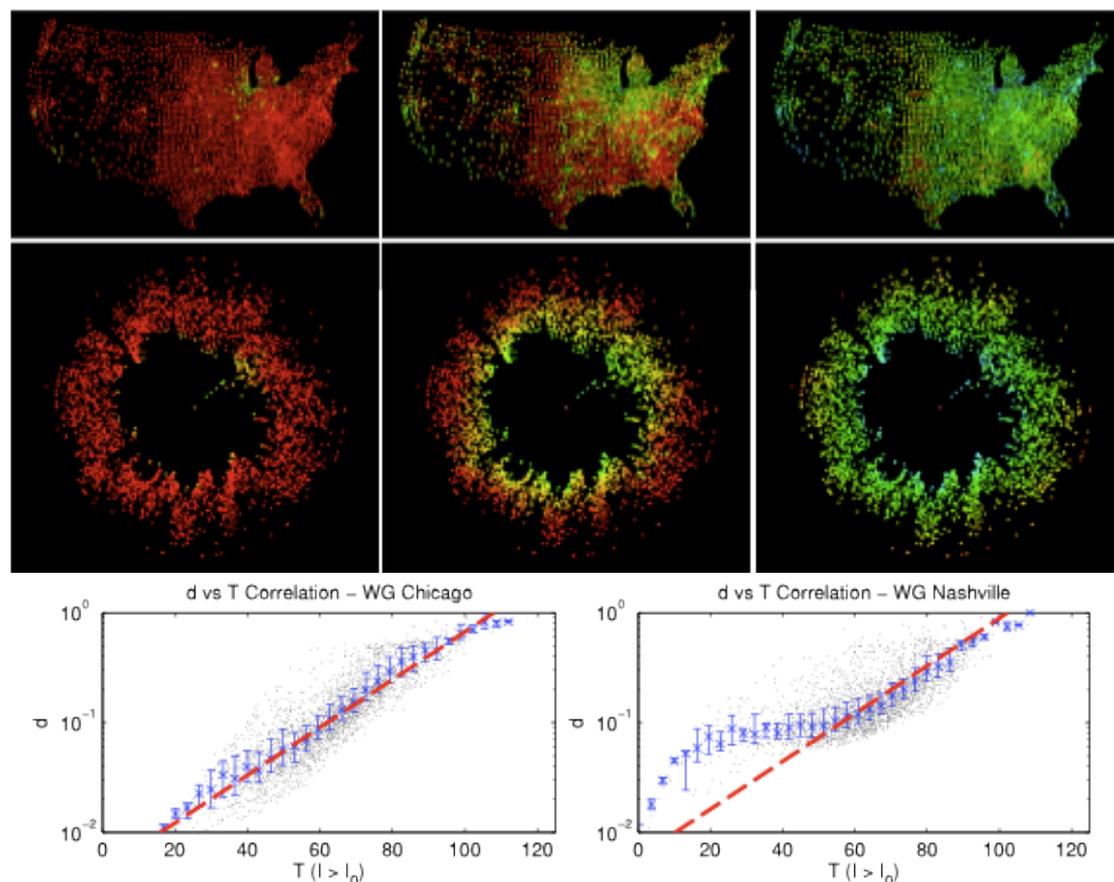


Figure 6: Global disease dynamics analyzed with shortest path tree tomography. The top row depicts a hypothetical disease spread in the United States. Geographically the pattern exhibits a high degree of complexity. Visualized in SPATO the pattern becomes symmetric and spreads concentrically. This representation permits to compute effective spreading speeds and determine how speeds related to system parameters. The analyses suggest that effective distance correlations directly with the arrival time of the epidemic. The lower panes depict effective speeds computed with this methods.

models on global disease dynamics often generate patterns that share many features irrespective of the specific modeling paradigm (stochastic metapopulation models, sophisticated and detailed agent based models, etc.).

The goal of this project was to investigate disease dynamics in the context of effective distances employed in the SPATO program described above. A very important insight of this project was that disease dynamics, irrespective of initial outbreak location exhibit characteristic symmetries in the effective distance representation. This is a direct consequence of the symmetries in the underlying transportation and mobility networks. Using epidemic arrival times and effective distances this approach permits the definition of effective spreading speeds despite the fact that in geographic representations a clear wave front cannot be determined when long range traffic is involved. Snapshots of disease dynamical processes in two representative mobility networks are illustrated in Fig. 6. Currently this project is being finalized and will be submitted for publication in the next few months.