Simulation and Analysis of Pathogen Transmission in an Agent- and GIS-based Model

Ryan C. Kennedy¹, Kelly E. Lane², S. M. Niaz Arifin¹, Hope Hollocher², Agustín Fuentes³, Gregory R. Madey¹

¹Department of Computer Science and Engineering
²Department of Biological Sciences
³Department of Anthropology

*This work supported in part by the National Science Foundation.
Overview

- Background
  - Basic macaque information
- Motivation
- Initial LiNK Model
- Improved LiNK Model
  - Performance Comparison
- LiNKStat – Social Network of Disease Transmission
- Future Work
Background

- Long-tailed Macaques
  - Coexisted with people across the island for at least 2000 years
  - Populations are female philopatric; males migrate from natal populations at age 7
  - Changing land use patterns have resulted in increased macaque-human interactions

- Disease Transmission
  - Primates are implicated as host and reservoir in several disease emergences
  - Increases in human to non-human primate interaction can lead to potential increases in bi-directional pathogen transmission
  - Long term human-macaque interface could become potential site of global disease emergence

(Macaca fascicularis)
Background

- Bali
  - Small Indonesian island
    (~130km × 80km)
  - Roughly 3 million people and 10,000 macaques
  - 42 Temple Sites
    - Majority of macaque population (and all females) reside here
    - Lots of macaque-human interaction at these sites
Where is Bali?
Biological/Anthropological Motivation

Research questions:
- What are potential rates and routes of pathogen transmission in macaques across the island?
- How do pathogen life history parameters impact this transmission?
- Do the answers change with the inclusion of humans as a component of the landscape?

- Landscape might play a very important role in each
  - Inclusion of GIS data
- Potential for pathogen transmission to humans!
Asian Monkey Temple Visitors Warned of Virus Transmission

SEATTLE, Washington, July 25, 2005 (ENS) - An international team of scientists concerned about transmission of viruses from monkeys to humans are warning travellers to Asia not to pet or feed the numerous monkeys they may see at temples and other tourist sites. Feeding the animals, or even carrying food into a temple, can greatly increase the risk of a bite or scratch, which can lead to transmission of infection.

The warning comes from researchers in the United States, Canada and Indonesia who have identified the first reported case in Asia of primate to human transmission of a retrovirus found in macaques and other primates known as the simian foamy virus. So far this virus has not been shown to cause disease in humans.

The transmission of the virus from a monkey to a human took place at a monkey temple in Bali, Indonesia, the researchers report in the July issue of the journal "Emerging Infectious Diseases."

Monkey at one of Bali's numerous monkey temples. New research shows these primates may carry viruses transmissible to humans. (Photo credit unknown)

Even though this particular virus may not prove dangerous to humans, the scientists warn that the dense human and primate populations in Asia could lead to other viruses carried by primates jumping the species barrier and causing human disease.

HIV, the virus that causes AIDS in humans, is believed to have originated as simian immunodeficiency virus, and jumped the species barrier to humans when African bushmeat hunters came into contact with blood from infected animals.

"The issue of primate-to-human viral transmission has been studied extensively in Africa, largely because that is where HIV originated," explains Dr. Lisa Jones-Engel, lead author of the study and a research scientist in the Division of International Programs at the University of Washington National Primate Research Center in Seattle.

"But there has not been much work on the topic in Asia, which has huge primate diversity and large human populations," she said.
Monkey Business!
More Monkey Business!

Due to the protection and resource availability found at these temple sites, macaques are able to exist in moderately high densities alongside high human densities.
Geography affects movement!
New Disease Emergences: Caused by Anthropogenic Changes to the Landscape

Many emerging infectious diseases of humans, domestic animals, and plants are driven to emerge by anthropogenic changes to the environment. The rapid growth of human populations in Southeast Asia, and subsequent increases in urbanization and development, has contributed to a rising level of human-wildlife interaction, leading to an ever-increasing source of zoonotic disease emergences. Non-human primates have been implicated as both pathogen host and reservoir in multiple disease emergences including SIV-HIV, SFV, Ebola virus, and Nipah virus, as well as several reemergent protozoa and helminth pathogens.
Role of Landscape

Traditional epidemiological models addressing pathogen transmission focus primarily on population-level patterns, i.e. the number of susceptible, infected, and recovered individuals of the SIR model, while neglecting the role of the landscape in transmission.
Human land-use patterns have resulted in a mosaic of riparian forest, small forest patches, agricultural lands, and urban areas across much of the island.
Macaque Sociology

- Macaques exist in large populations in “monkey forests” within each temple complex consisting of between 30 and 350 individuals, composed predominantly of between one and five matrifocal groups.

- The matrifocal society of the macaques results in strong female philopatry.

- Male macaques, in dispersing, may be acting as both units of gene flow between seemingly isolated macaque populations as well as a mechanism of pathogen transmission across the island.
LiNK Model Background

- Initial version was developed and presented at SwarmFest 2006
  - Developed by multiple programmers
  - Coded without a lot of prior knowledge or experience in modeling with GIS data
  - Utilizes raster files for movement behavior
  - Worked well, but numerous possible enhancements
    - Didn’t scale as well as hoped
  - Movement
    - Probability-based
    - Can be based upon arbitrary number of cells outward
LiNK Model Background

- Several Problems
  - Computational overhead because of large raster files
    - Macaque movement dependent on queries to each raster file
  - Working heavily with raster files can be slow (in our application)
  - Inconsistency between pathogen parameters within and outside of temples
  - Tweaking parameters required biologists to manually edit java code
  - Huge output sets representing “social networks” of disease transmission – difficult to analyze

- Goal: Use lessons learned from initial model to develop more intelligent, efficient, and consistent code and an overall better model
Related Work

- Numerous studies with ABM combined with GIS
- Very few studies with ABM and GIS in which agents continually interact with complex GIS data (but some papers/posters at NAACSO 2009!)
  - Keeling’s model of the spread of foot-and-mouth disease is most notable, but there is no movement in the model
- Very few ABM models of pathogen transmission with dynamic interaction with GIS data on a large scale
LiNK Model

- Agents
  - Macaques with studied behavior and habitat probabilities

- Environment
  - GIS data (ESRI shapefiles) for 7 type of environment + 3 user-generated shapefiles

- Model Parameters
  - Pathogen
  - Movement
  - Behavior

- Integration of Tools
  - Simulation Environment
    - Repast
  - Display
    - OpenMap, Repast
  - Spatial Information
    - GeoTools
    - JTS Topology Suite

- 12 simulated hours per timestep
Parameters

Clearance Time

Latency  Symptomatic  Acquired Immunity

Infection  Infection Free  Susceptible

Time

Latent Infection

Susceptible  Symptomatic Infection

Acquired Immunity
Possible Transmission Cycle

- $M_1$ infectious, $M_2$ susceptible
- $M_1$ will infect $M_2$ based on virulence and infectivity
- If $M_1$ infects $M_2$ …
Possible Transmission Cycle

- $M_2$ is now infectious and asymptomatic
- Once latency period passes …
Possible Transmission Cycle

- $M_2$ is now symptomatic
- More time passes …
Possible Transmission Cycle

- M₁ has clears infection and has an acquired immunity
- M₂ still infectious
GIS Improvements
GIS Data

- Very detailed GIS data for Bali
  - Forests, lakes, rivers, rice fields, roads, cities
- Shapefile format

- Continuously interacting with complex GIS data is difficult for thousands of agents
Raster vs. Vector Representation

Raster

Vector
Raster Files

- Shapes are represented in a grid-like manner
  - Pixels

- Advantages
  - Fairly easy to determine if you are “on” a certain feature
    - Either on or off

- Disadvantages
  - Requires large matrix to represent features
    - Lots of overhead
Vector Files

- Shapes are represented by coordinates (points, lines, polygons)
  - Paths

- Advantages
  - Small file size
    - Less overhead
  - GIS files collected as coordinates
    - Less data loss
  - Math-based queries

- Disadvantages
  - Query time to determine if you are in a “complex” feature is lengthy
    - Often must approximate some data
Spatial Approximation

- Data Approximation
  10,000 data points $\rightarrow$ 100 data points
Precalculated Spatial Query

- Use spatial query tools to build a simple matrix to represent each shapefile
  - Build once (at a high cost), at a specified granularity (currently about 100 meter increments)
  - Stored in arrays and written to file
  - Java Serialization is utilized to reload data
  - Still use probability-based movement

- Bottom line:
  - More macaques can do more things more quickly
## Performance Comparison

### GUI Load Time (s)

<table>
<thead>
<tr>
<th></th>
<th>Coast, Lakes</th>
<th>Coast, Lakes, Forests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Query</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Raster Query</td>
<td>35</td>
<td>42</td>
</tr>
<tr>
<td>Simplified Spatial Query</td>
<td>1.8</td>
<td>2.5</td>
</tr>
<tr>
<td>Precalculated Query Matrix</td>
<td>1.6</td>
<td>2</td>
</tr>
</tbody>
</table>

### Timesteps/sec

<table>
<thead>
<tr>
<th></th>
<th>Coast, Lakes</th>
<th>Coast, Lakes, Forests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Query</td>
<td>1.6</td>
<td>0.15</td>
</tr>
<tr>
<td>Raster Query</td>
<td>18.5 (11x faster)</td>
<td>19 (126x)</td>
</tr>
<tr>
<td>Simplified Spatial Query</td>
<td>39.5 (25x)</td>
<td>15.8 (105x)</td>
</tr>
<tr>
<td>Precalculated Query Matrix</td>
<td>126.2 (79x)</td>
<td>124.1 (827x)</td>
</tr>
<tr>
<td>Precalculated Query Matrix, non-GUI</td>
<td>669.6 (419x)</td>
<td>650.2 (4335x)</td>
</tr>
</tbody>
</table>
Performance Comparison

![Bar chart showing performance comparison between different methods and their combinations. The x-axis represents different methods: Spatial, Raster, Simplified Spatial, Precalculated Query Matrix, Precalculated Query Matrix, non-GUI. The y-axis represents Timesteps/s, ranging from 0.1 to 1000. Each bar is divided into two sections, one for "Coast, Lakes" and the other for "Coast, Lakes, Forests." The chart illustrates a significant improvement in performance for the Precalculated Query Matrix methods compared to the others.]
## Performance Comparison

<table>
<thead>
<tr>
<th>Number of Initial Dispersed Macaques</th>
<th>10</th>
<th>100</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raster Query (3 Layers)</td>
<td>51.3</td>
<td>29</td>
<td>19.9</td>
</tr>
<tr>
<td>Raster Query (7 Layers)</td>
<td>33.6</td>
<td>27.6</td>
<td>11</td>
</tr>
<tr>
<td>Precalculated Query Matrix (3 Layers)</td>
<td>140.7</td>
<td>131.4</td>
<td>83.8</td>
</tr>
<tr>
<td>Precalculated Query Matrix (7 Layers)</td>
<td>137.5</td>
<td>129.5</td>
<td>82.9</td>
</tr>
<tr>
<td>Precalculated Query Matrix, non-GUI (3 Layers)</td>
<td>669.8</td>
<td>487.8</td>
<td>154</td>
</tr>
<tr>
<td>Precalculated Query Matrix, non-GUI (7 Layers)</td>
<td>680.4</td>
<td>529.6</td>
<td>158.2</td>
</tr>
</tbody>
</table>
Scalability

![Graph showing scalability with number of initial dispersed macaques vs. timesteps. Legend includes different colored lines and points representing various query methods with different layer counts and GUI options.]
## Recommendations

<table>
<thead>
<tr>
<th></th>
<th>Raster Query</th>
<th>Spatial Query</th>
<th>Simplified Spatial Query</th>
<th>Precalculated Query Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy of GIS Data</strong></td>
<td>Good</td>
<td>Excellent</td>
<td>Very Good</td>
<td>Very Good</td>
</tr>
<tr>
<td><strong>Amount of GIS Data</strong></td>
<td>Good</td>
<td>Poor</td>
<td>Fair</td>
<td>Excellent</td>
</tr>
<tr>
<td><strong>Complexity of GIS Data</strong></td>
<td>Fair</td>
<td>Excellent</td>
<td>Very Good</td>
<td>Very Good</td>
</tr>
<tr>
<td><strong>Load Time</strong></td>
<td>Poor</td>
<td>Very Good</td>
<td>Very Good</td>
<td>Excellent</td>
</tr>
<tr>
<td><strong>Memory Requirement</strong></td>
<td>Fair</td>
<td>Very Good</td>
<td>Very Good</td>
<td>Excellent</td>
</tr>
<tr>
<td><strong>Number of Agents</strong></td>
<td>Very Good</td>
<td>Poor</td>
<td>Fair</td>
<td>Very Good</td>
</tr>
<tr>
<td><strong>Timesteps/s</strong></td>
<td>Very Good</td>
<td>Poor</td>
<td>Fair</td>
<td>Excellent</td>
</tr>
</tbody>
</table>
LinkStat

Social Networks Analysis of Disease Transmission

UNIVERSITY OF NOTRE DAME
LiNKStat

- LiNK is complex and can create enormous amounts of data
  - LiNKStat was developed to glean scientific insight and validation on LiNK

- LiNKStat
  - Written in perl with Tk
  - Interactive, graphical display
  - Ability to track route of infection over time
    - The “social network” of disease transmission
    - Number of macaques directly and indirectly infected over time
  - Birth, death, and infection events are reported
  - General statistics
LiNKStat Interface

<table>
<thead>
<tr>
<th>Infector: 99.354.1</th>
<th>Number of Infection Event(s): 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timestamp</strong></td>
<td><strong>Infection</strong></td>
</tr>
<tr>
<td>6</td>
<td>99.142.1</td>
</tr>
<tr>
<td>9</td>
<td>99.354.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infector: 99.355.1</th>
<th>Number of Infection Event(s): 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timestamp</strong></td>
<td><strong>Infection</strong></td>
</tr>
<tr>
<td>2</td>
<td>99.424.1</td>
</tr>
<tr>
<td>3</td>
<td>99.355.1</td>
</tr>
<tr>
<td>9</td>
<td>99.905.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infector: 99.356.1</th>
<th>Number of Infection Event(s): 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timestamp</strong></td>
<td><strong>Infection</strong></td>
</tr>
<tr>
<td>14</td>
<td>99.356.1</td>
</tr>
</tbody>
</table>
LiNKStat Pathogen Transmission Graph
Average number of infections, both total number of infection events and total number of temples infected, when all GIS layers were available for macaque decision making and when only the coast GIS layer was available for macaque decision making.
Average number of total infections and total temples infected with only the coastline GIS layer, with all layers, and with each layer cycled off independently. Note the substantial disparity between iterations with only the coast available and with all layers available as well as the increase in infection when the urban layers (road and city) were left off.
Model Building Results

- We demonstrate the importance of integrating GIS landscape data with biological data to help understand and predict patterns at the population level.
- Our model provides predictive power to identify patterns that can be more closely examined in the field.
- This model will likely support the hypothesis that male dispersal acts as a genetic link between populations island-wide and will act as a mechanism in disease transmission.
- Model results will have potential long term conservation implications for macaque survival.
- Enhancements of interaction of GIS data and agents have allowed us to develop a more complex and scalable model.
  - LiNK can run in GUI mode on any computer with Java or on a high-performance computing cluster with the GUI disabled.
LiNK Summary

- Uses Java serialization to load precalculated matrices
  - Influence behavior appropriately
  - Can easily access surrounding landscape data
- Uses vector shapefiles for background display only
- GUI can be disabled for faster simulations
- Runs without GUI on a cluster
- Uses Perl serialization to load LiNKStat graphs
  - Decreases load time of LiNKStat
Future Work

- Deploy website front-end
- Allow for easy inclusion of custom GIS data
- Allow for more computationally complex models
  - Inclusion of human agents
  - Inclusion of more complex pathogens
Questions or Comments?

UNIVERSITY OF NOTRE DAME