

appear in the printed materials of the Symposium and also archived online.

These summaries allow you to give a general overview of your tools, significantly highlight novel features, provide references to papers and other relevant work and describe any new discoveries you made about your tools while working through the Challenge problem. Only the two-page summaries of the best entries (which are awarded an award) will be published in the Proceedings. Nevertheless, ALL submitted two-page summaries will be published online - along with your answer - in the VAST Benchmark Repository, whether or not they received an award.

The two-page summary should be formatted according to the general IEEE VGTC Guidelines

<http://www.cs.sfu.ca/~vis/Tasks/camera.html>

MINI CHALLENGE 1 : Text Records - Investigations into Arms Dealing

You have been charged with providing a forensic analysis of illegal arms dealing that may have taken place last year. The following contains intelligence reports and material drawn from other sources that will be of concern to you in your analysis. The information begins in 2008, and runs through early 2009. Please organize your analysis by country. In particular, we are interested in the following:

MC1.1 Summarize the activities that happened in each country with respect to illegal arms deals based on a synthesis of the information from the different report types and sources. State the situation in each country at the end of the period (i.e. the end of the information you have been given) with respect to illegal arms deals being pursued. Present a hypothesis about the next activities you expect to take place, with respect to the people, groups, and countries.

Provide a Detailed Answer.

MC1.2 Illustrate the associations among the players in the arms dealing through a social network. If there are linkages among countries, please highlight these as well in the social network. Our analysts are interested in seeing different views of the social network that might help them in counterintelligence activities (people, places, activities, communication patterns that are key to the network).

Provide a Detailed Answer.

Also for each mini challenge entry provide:

- a Video showing how you conducted the analysis
- the optional Two-Page Summary

MINI CHALLENGE 2 : Hospitalization Records - Characterization of Pandemic Spread

There was a major epidemic outbreak that spanned several cities across the world in 2009. The disease tended to move fast and be fairly difficult to combat. Health officials are seeking help to analyze the illness across these countries to help characterize the spread of the disease.

You have been provided with hospital admittance and death records for cities involved in the epidemic. Please examine the following:

MC2.1: Analyze the records you have been given to characterize the spread of the disease. You should take into consideration symptoms of the disease, mortality rates, temporal patterns of the onset, peak and recovery of the disease. Health officials hope that whatever tools are developed to analyze this data might be available for the next epidemic outbreak. They are looking for visualization tools that will save them analysis time so they can react quickly.

Provide a Detailed Answer.

MC2.2: Compare the outbreak across cities. Factors to consider include timing of outbreaks, numbers of people infected and recovery ability of the individual cities. Identify any anomalies you found.

Provide a Detailed Answer.

Also for each mini challenge entry provide:

- a Video showing how you conducted the analysis
- the optional Two-Page Summary

MINI CHALLENGE 3: Genetic Sequences – Tracing the Mutations of a Disease

this case concerns a patient, identified by Interpol as Nicolai Kuryakin, admitted to a hospital in Paris after being removed from a flight to Moscow due to illness. The patient, now deceased, was admitted with an unidentified illness and later developed symptoms consistent with Drafa Fever. An autopsy confirmed the presence of the Drafa virus in the patient's bloodstream.

Investigators need your help to understand the evolution of the current Drafa virus outbreak, as it may shed some light on Nicolai's contacts. In question 1 and 2, you need to characterize the distance between genetic sequences and reconstruct the likely evolutionary paths. Public health organizations also need your help to understand how characteristics of the disease relate to the new virus strains. In particular, they would like to know what changes in genetic sequence make the symptoms more severe (question 3) and what changes make the virus most dangerous overall (question 4).

Technical Background on Viruses and Mutation

When the Drafa virus infects a host, it uses the host's cell machinery to replicate and evolve. When the virus replicates, it creates viral particles consisting of protein molecules. In this way, the virus makes copies of itself, growing the population of virus particles within the same host and eventually spreading to others. The proteins that make up these particles affect the ability of the virus to create disease, evade host defense mechanisms, and protect its genetic information. This genetic information includes the genes which can be thought of as the instructions on how to create the various viral protein molecules. Viruses can have very few to hundreds of such genes.

Since the Drafa virus has a high mutation rate, the genetic information contained within the virus is constantly changing. These mutations occur as a result of mistakes made during the viral replication process. The genetic code (also known as sequence) gets rewritten over time, resulting in a stronger or weaker virus. Each of these variations, if viable, creates a new viral strain or mutant.

A gene consists of a sequence of single molecules (each called a base or a nucleotide) and coded as A, T, C, or G. Triplets of these bases combine to form amino acids (for example GGC is Glycine and CAA is Glutamine) and amino acids combine to form proteins.

For this scenario, we are focusing on the genetic sequence information for a single Drafa gene, the one that codes for a surface protein on the virus. The location of this protein on the surface enables portions of it to interact with other protein molecules in the host. As such, evolution of this protein greatly affects the virulence of Drafa (the Drafa virus), its relative ability to cause disease.

As the Drafa virus spreads from host to host, it mutates and evolves. Each mutation substitutes an existing base with a different one (for example, A changing to C, G changing to C, ...). Over time, this results in evolved viral strains (mutants) with a number of modified bases.

For the given scenario, diagnostic tests confirm that the current outbreak is a new evolved form of the Drafa virus. Viral mutants have been isolated from human blood samples in hospitals where patients are being treated, and laboratory sequencing techniques are being used to collect genetic information.

MC3.1 What is the region or country of origin for the current outbreak? Please provide your answer as the name of the native viral strain along with a brief explanation.

Provide a Short Answer.

MC3.2 Over time, the virus spreads and the diversity of the virus increases as it mutates. Two patients infected with the Drafa virus are in the same hospital as Nicolai. Nicolai has a strain identified by sequence 583. One patient has a strain identified by sequence 123 and the other has a strain identified by sequence 51. Assume only a single viral strain is in each patient.

Which patient likely contracted the illness from Nicolai and why? Please provide your answer as the sequence number along with a brief explanation.

Provide a Short Answer.

MC3.3 Signs and symptoms of the Drafa virus are varied and humans react differently to infection. Some mutant strains from the current outbreak have been reported as being worse than others for the patients that come in contact with them.

Identify the top 3 mutations that lead to an increase in symptom severity (a disease characteristic). The mutations involve one or more base substitutions. For this question, the biological properties of the underlying amino acid sequence patterns are not significant in determining disease characteristics.

For each mutation provide the base substitutions and their position in the sequence (left to right) where the base substitutions occurred. For example,

C → G, 456 (C changed to G at position 456)

G → A, 513 and T → A, 907 (G changed to A at position 513 and T changed to A at position 907)

A → G, 39 (A changed to G at position 39)

Provide a Short Answer.

MC3.4 Due to the rapid spread of the virus and limited resources, medical personnel would like to focus on treatments and quarantine procedures for the worst of the mutant strains from the current outbreak, not just symptoms as in the previous question. To find the most dangerous viral mutants, experts are monitoring multiple disease characteristics.

Consider each virulence and drug resistance characteristic as equally important. Identify the top 3 mutations that lead to the most dangerous viral strains. The mutations involve one or more base substitutions. In a worst case scenario, a very dangerous strain could cause severe symptoms, have high mortality, cause major complications, exhibit resistance to anti viral drugs, and target high risk groups. For this question, the biological properties of the underlying amino acid sequence patterns are not significant in determining disease characteristics.

For each mutation provide the base substitutions and their position in the sequence (left to right) where the base substitutions occurred. For example,

C → G, 456 (C changed to G at position 456)

G → A, 513 and T → A, 907 (G changed to A at position 513 and T changed to A at position 907)

A → G, 39 (A changed to G at position 39)

Provide a Detailed Answer.

Also for each mini entry provide:

- a Video showing how you conducted the analysis
- the optional Two-Page Summary

Grand Challenge: Arms Dealing and Pandemic

In Mini-Challenge 1, you investigated illegal arms deals that involved several countries. In Mini-Challenge 2, you investigated a pandemic outbreak of a virus across several cities in the world. In Mini-Challenge 3, you investigated the source of a virus strain taken from a victim of the pandemic.

For the Grand Challenge, you are charged with investigating any possible linkage between the illegal arms dealing and the pandemic outbreak.

In particular, you need to address the following:

1. Briefly describe your hypothesized linkage between the arms dealing activity and the pandemic outbreak.
2. Where did the disease originate (as far as you can tell with the data)? Provide a time-based tracing of its spread among countries. If you have established linkages between arms dealers in certain countries and the pandemic, please indicate this.
3. We had countries with arms dealers identified in MC 1 that did not suffer pandemic outbreaks in MC 2. Provide a hypothesis as to why some countries that may have been involved with arms dealers did not suffer an outbreak?

Recall that we are particularly interested in how visualizations helped you with your analysis. Remember to explain in your process what visualizations you used and what insights were gained from these visualizations.

No additional data set is needed for the Grand Challenge (only the data provided in mini 1, 2 and 3).

For the grand challenge you need to provide:

- a debrief
- a Video showing how you conducted the analysis
- the optional Two-Page Summary