Vol 8. No. 1 Issue 2 – May, 2015
African Journal of Computing & ICT

www.ajocict.net

Immunization Model for Ebola Virus in Rural Sierra-Leone

A.A. Ojugo., D. Allenotor. & D.A. Oyemade.
Department of Mathematics/Computer Science
Federal University of Petroleum Resources
Effurun, Delta State, Nigeria
ojugo.arnold@fupre.edu.ng, allenotor.david@fupre.edu.ng, oyemade.david@fupre.edu

R.E. Yoro
Department of Computer Science
Delta State Polytechnic
Ogwashi-Uku, Delta State
rumerisky@yahoo.com

C.N. Anujeonye
Dept of Computer Education
Federal College of Education Technical
Asaba, Delta State, Nigeria
anujeonyechristy@gmail.com

ABSTRACT

Declared an epidemic for public health emergency of International concern in 2014 by World Health Organization (WHO), Ebola Virus Disease was first noticed in Guinea. Many studies since then has ensued to provide various safe plans that implemented, will contain the sporadic spread of the outbreak. Relations among individuals in a society help them achieve life's goals. With the society as a graph, diffusion helps spread ideas as society evolves into a complex network. EVD spread via diffusion is modeled as complex design as we seek solution to processes not limited to spread cum propagation management, immunization and capacity service distribution etc. Our study uses the stochastic SI-dynamic models to predict EVD-spread effects using Moyamba District in Sierra Leone (as dataset) to address immunization task via computation of the expected spread and spread minimization.

Keywords — Stochastic, immunize, network, vertices, SIS, SIR,

African Journal of Computing & ICT Reference Format:
A.A. Ojugo., D. Allenotor, D.A. Oyemade, R.E. Yoro & C.N. Anujeonye (2015): Immunization Model for Ebola Virus in Rural Sierra-Leone. Afr J. of Comp & ICTs. Vol 8, No. 1, Issue 1. Pp 1-10.

1. INTRODUCTION

EBOLA Virus Disease / Hemorrhagic Fever is one of many severe, viral hemorrhagic fevers [33] with high fatality rate as in its subjects (humans and non-humans). It infection is caused by the Filiviridae family such as the Genus Ebolavirus. Its symptoms manifest between a period of 2 and 21-days – with incubation period of 8-10days, and results in sporadic outbreak and fatality rates of over 90percent. It has been found rampant in rural areas and tropical rainforests. Its original host is traced to Fruit *Bats* (Pteropodidae), has that the Genus Ebolavirus is 1-of-3 members of Filoviridae family. Others are Genus Cuevavirus and Genus Marburgvirus. The Genus Ebolavirus has 5-different species: (a) Bundibugyo Ebolavirus (BDBV), (b) Zaire Ebolavirus (EBOV), (c) Reston Ebolavirus (RESTV), (d) Sudan Ebolavirus (SUDV) and (e) Tai Forest Ebolavirus (TAFV) [33].

1.1 Nature and Characteristics of EVD

The nature of EVD (especially for the Genus Filoviridae) is such that its sporadic outbreaks is not mainly restricted to Africa as there are also confirmed cases in the United States of America, United Kingdom and other West African regions – where the virus is currently ravaging nations such as Sierra Leone, Ivory Coast, Sudan, Gabon, Uganda etc and with only Nigeria, whom seemed to have managed the situation with success stories in her many states. Most isolated outbreaks are unnoticed – as the outbreak has characteristic feats to appear in mostly health settings and amongst health worker who have tried and given their lives to contain the situation and sporadic outbreak [33].

1.2 Spread and Symptoms

Though there are reported cases of fully recovered from the EVD – the World Health Organization (WHO) has also noted that patients who have made such recovery can still transmit as they are still virus carrier for up to 7-weeks after such full recovery. Laboratory findings include the presence of low white blood cells, platelet count and elevated liver enzymes. Its spread has been found to be via direct contact with an infected person or animal, blood or secretion from infected person, contact with contaminated medical equipment, reusing unsterilized needles, eating carcass of infected animal, inhaling contaminated air in hospital environment and non-implementation/observance of the universal precaution. Its many symptoms includes fever, headache, diarrhea, constant vomiting, chest pain, sore throat, stomach pain, cold, cough, joint and muscle pains, weakness, depression, confusion, red eyes as well as internal and external bleeding. It is diagnosed from several tests that include:

a. Enzyme Linked immune-Sorbent Assay (ELISA)
b. Antigen Detection Tests
c. Serum Neutralization Test
d. Reverse Transcriptase Polymerase Chain Reaction
e. Virus Isolation via Cell culture

In diagnosing EVD, other diseases that must be isolated from it includes malaria, typhoid fever, plague, shigellosis, cholera, leptospirosis, hepatits, meningitis, relapse fever, rickettsiosis amongst other viral haemorrhagic fevers.

1.2 Prevention of EVD

EVD can be prevented thus: (a) maintain good hygiene and sanitation practices, (b) isolate (confirmed) infected patients, (c) sanitize burial rites by properly disposing dead bodies of Ebola confirmed patients in safe manner, (d) use disposable needles and avoid its reuse, and (e) use of personal protective equipment such as gloves and use of gowns always.

Fig 1: EVD Spread at time t = 0

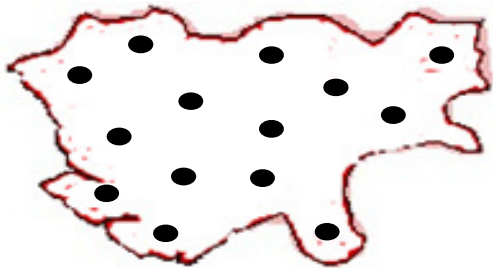

Fig 2: EVD Spread at time t = 1

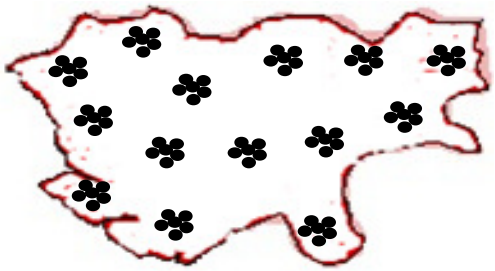

Fig 3: EVD Spread at time t = 3

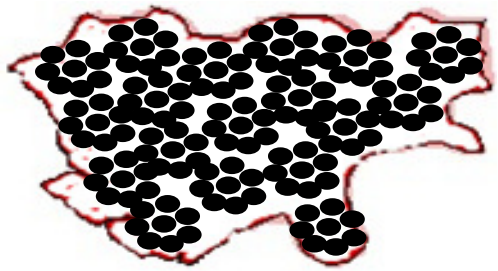

Fig 4: EVD Spread at time t = 4

2. THE ACTOR-BASED NETWORK/GRAPH MODEL

Networks consists of links and/or that bind nodes together through a predefined network model so that the research can analyze network entities along various theories to explain the observed patterns within [11, 12]. Thus, network helps propagate local feats present in nodes that eventually emerge as global patterns. It examine dynamics in relationship between nodes as well as helps to locate all the influential entities within such a network – as it theoretically, allows connection convergence of nodes [13].

Visual representation of such networks helps convey results of explored nodal-links and complex feats captured via signed structural quantities that represent node relationships. Positive edges for positive relations (marriage, friendship, colleagues etc); while negative edges are for negative relation (hatred, anger etc). If A→B and B→C are positive; C→A is a negative – it is an *unbalanced* cycle. Also, if B→A is positive, and both A→C and B→C are negative – these groups will be likely morphed into *balanced* one. Such concept is used to predict tie-behavioural evolution for both *balanced* (where actors will unlikely change their opinions about other actors in same

group) and an *unbalanced* (actors will likely change opinions of others) cycles [29, 10].

A powerful role of networks is to bridge local feats as it blossoms into global patterns – explaining how simple node and its links impacts a complex effect that ripple through a population system. Each node shapes the graph's evolution in time, adapting themselves to varying forms as need arises, which defines tie-strengths. Graph is denoted as G = (V,E). Each node $i \in V$ has set of ties $m \in E$ that is either self-linked (loop), single- or multi-link (undirected/directed). Each node has a corresponding set of actors to which they are either linked or isolated from. The links are either weak or strong to note the social relationship between such nodes, which are measured via dyads *D*.

A graph or networks holds these definitions as true:

1. Node States – a node may be in any of: (i) *susceptible*: if the node is not exposed but will soon be, (ii) *active*: if node is been exposed and has adopted the innovation as well as influence others to do same, and (iii) *removed*: if the node had been exposed to corruption and has been recovered.{1}[2}
2. A graph (directed/undirected) has nodes as actors ($v \in V$) with ties $(u,v) \in E$ that allows interaction between the nodes. Graphs are drawn from a specific family (and no algorithm considers all possible graphs). If G = (V,E) is a dynamic network, E its set of edges that are time-stamped, $(u,v)_t \in E$ are interactions at time $t \in Z^+$. If an innovation (data) is introduced into the network, a node can assume any one of 3-states as above. In a graph's typical setting and for a diffusion model, nodes are initialized as *active* – so that in discrete time-steps and at each time evolution, an active actor/node is exposed to the innovation for adoption. It continues till a stop criterion is satisfied or there are no more inactive nodes.
3. Knowledge of the propagation model will help us plant *d* copies of seed-set nodes that have adopted or is exposed the EVD in the network so as to maximize speed of its spread given by F_d. An adaptive seed-set has knowledge of choices made by the immunization algorithm; while a randomized seed-set simply places uniform cum random copies of seed-set on the network. Virus propagation model that determines how the virus is spread on graph.
4. Immunization model aims to minimize viruses spread and an immunized node cannot transfer/receive a virus. It is conceptually removed from graph. Cost of immunization model is, number of nodes immunized.
5. Adversary with knowledge of propagation model, plants *d* copies of malware in network so as to maximize speed of its spread is denoted as F_a. An adaptive adversary is one who has knowledge of choices made by the immunization algorithm; while a randomized adversary places copies of virus, uniformly at random on the network.
6. Ties – links that connects various nodes together. A tie continuum expresses relationship spectrum as provided by *tie strengths.* Whether a link between nodes exists or not, relationships have their own properties that defines the tie-strength between such nodes/agents as they form clusters, cliques and communities [20]. Thus, ties are measured via *Dyads* (pair of interactions between *n*-actors with *m*-ties that result in an m×n binary matrix of elements in G. Many have embarked on social graphs to measure tie-strength as a linear combiner with the potential of a feedback into the society in ways that benefit a user via Eq. 3 given by TS_i to represents number of predictive variables used in task, e_i is error term, D_i is dyads pairs, N_i is network structure and EI_i are external influences.

$$TS_i = \alpha + \beta R_i + \gamma D_i + N_i + EI_i + e_i \quad (3)$$

7. Network Structure – A node has ties with other nodes that helps it form clusters. The ties are based on strengths as relationships that ensue between the said node and other nodes. A network is made up of clusters (strong candidates of cohesive cliques). As cluster increases, the nodes forms more cliques with others in close proximity so that highly clustered network have short-path length with cluster coefficient that influences the number of nodes that are exposed at any given time. With such EVD contagion, the network experiences reinforced adoption as more nodes are likely to be exposed via overlapping influences. Thus, we have that N_i encodes an idea that its structure is dependent on structural predictive variable as regards: (a) tie strength, (b) reciprocity, and (c) cohesive clusters/its coefficients etc as given by Eq. 4 [3] and N_i uses the parameters: (a) $P_{\theta,G}$ to encode the graph's probability distribution of nodes via defined upper and lower bound of agent dispositions in the system, (b) μ and Med are node/network threshold, (c) $\sum_{t=0}^{t-1}\sum_{i\in L}\lambda_t(s-\mu_L)^t$ is learning with momentum and function of outcome convergence in time, and (d) Max/Min are the upper/lower bounds of final adopters.

$$N_i(G) = P_{\theta,G} + \lambda_\mu \mu_L + \lambda_{Med} Med_L + \sum_{t=0}^{t-1}\sum_{i\in L}\lambda_t(s-\mu_L)^t + \lambda_{Min} Min_L + \lambda_{Max} Max_L \quad (4)$$

8. Immunization model aims to minimize spread of EVD so that more nodes stay immunized and cannot engage in propagation ever. Such nodes are conceptually removed from the graph network. Cost of immunization model is given by the number of nodes immunized.

1.3 Graph Network Types

Social graphs are tools best suited for diffusion of an idea or innovation. It helps spread of data – making it easier for users to disseminate useful data as well as viruses. The problem of virus propagation has been a recurring subject and ongoing research notes that every harmful data spread over such networks are considered as malware or viruses as can be interchangeably used; while the process of impeding the spread of such harmful data (malware) over such social network is referred to as network immunization. This aims to prevent the spread of such malwares, protect such networks from virus attacks and control data and sensitive information leakages – while at same time noting that our resources such

www.ajocict.net

as vaccination and antiviral influences are costly and limited in their capability to discover such malware. With such AVs and vaccinations, users aim to achieve the best effect; while still allocating the least resources possible [22, 23, 24].

Hackers (or adversary) wreak more havoc being aware of the propagation model used to avert such attacks. In simplest form, a social network is seen as a complex graph. Thus, the propagation model has as input a graph G = (V,E), state vector $S_v^{(t)}$ for each node vertex v ∈ V at t, and an internal parameter vector P. Based on the states of all interacting nodes, it outputs a new state vector $S_v^{(t+1)}$ for each node at t+1 (Giakkoupis et al, 2010). Models are applied to synthetic dataset with graph types [19, 17, 25]:

a. **Scale-Free Networks:** Probability that node *x* in network is of degree k is proportional to $k^{-\gamma}$ with $\gamma > 1$. Scale free graph are modeled as by Barabasi and Albert [4]. It inserts nodes sequentially with each node linked to an existing one chosen with a probability proportional to its current degree in a tree-fashion with grandparent-parent-children-grandchildren structure and it builds graph with exponent $\gamma = 3$ denoted with G_{sf}. Each node in the graph can be autonomous but must be connected to an existing one. Thus, two nodes are connected together on the graph via physical link between two corresponding autonomous systems. Such is referred to as Autonomous Systems. Another scale free graph consists of undirected edges between nodes, also termed Co-Author graphs.
b. **Small World Graphs** are those with small characteristics path length L (the average shortest path between any pair of vertices) and large clustering coefficient C (the average fraction of pairs of neighbours of a node also connected to each other). We use Watts [32] to generate small-world graph denoted as G_{swL} with path length feat; while G_{swC} to denote those of large clustering coefficient. Graphs of G_{swL} are influenced by α that intuitively determines the probability of two nodes being connected given a number of their common neighbours. It controls the extent a graph has small/densely connected components so that as α nears infinity, G_{swL} becomes a random graph. Conversely, G_{swC} is influenced by q, which determines the probability of an edge in the lattice being rewired to connect to a random node in the graph. Thus, initialized on a ring lattice, each node is of degree k. Small values of q entails G has high clustering coefficient and large average path length; while large values q creates random graphs. For q-values close to 0.01, the generated graphs are small-world graphs. Also G_{swL}, G_{swC} and G_{sf} are quite distinct graphs.

2. SUSCEPTIBLE INFECT MODELS FOR EPIDEMIC SPREAD

The two major susceptible-infect models Susceptible-Infect-Remove and Susceptible-Infect-Susceptible. In SIR, a node may be in any of these states: (a) susceptible: if node has no virus but will become infected if exposed to it, (b) infected: if node has the virus and can pass it to others, and (c) removed: if node had the virus but has been recovered or virus dies. The node is permanently immunized and can no longer participate in propagation, and a particular node cannot be infected twice. Conversely in SIS, a node can be cured but not immunized. Thus, it can be infected again. Such node switches between susceptible and immunized.

2.1 Independent Cascade Model

It is a discrete-time special case SIR model in which at t = 0, an adversary inserts *d* copies of virus to some nodes on graph. If node *x* is infected the first time at t, it has single chance to infect any neighbours *y* currently uninfected. Probability that *x* succeeds with *y* is P_{xy}. If *x* succeeds, *y* is infected at t+1; Else, *x* tries again in the future (even if *y* gets infected by another neighbour). This process continues and stops after *n*-steps if no more infections are possible. It requires a nodes stay infected exactly once and it is the independent cascade model following [18]. Graph of size M, has M_d subset of nodes and *d* copies of virus placed on the network. With propagation complete, $S(M_d,G)$ is expected number of infected nodes. Expectation exceeds all random choices made by propagation model. Eq. 1 is maximum expected number of infected nodes and maximum exceeds all possible initial virus placements.

$$S_d(G) = \max_{M_d} S(M_j, G) \quad (1)$$

Subset $A_d = argmax_{M_d} S(M_d, G)$ corresponds to choices made by an adaptive adversary. $S_d(G)$ is epidemic spread in G and a similar definition of epidemic spread of randomize adversary as in Eq. 2 in which case, define it to be the expected epidemic spread where the expectation takes over all possible positions of the *d* viruses placed on the network and given by:

$$S'_d(G) = E_{M_d}[S(M_d, G)] \quad (2)$$

2.2 Dynamic Propagation Model

In SIS, viruses are seen as dynamic birth-death process that evolves overtime. It continues to either propagate or eventually die. An infected node *x* spreads virus to node *y* in time t with infection rate of $\frac{\beta}{\delta}$ and probability β. At same time, an infected node may recover with probability δ. With adjacency matrix T, $\lambda_1(T)$ is largest eigen-value of T. The condition $\frac{\beta}{\delta} < \frac{1}{\lambda_1(T)}$ holds true as epidemic threshold and is sufficient for quick recovery, easily proven (Ganesh et al, 2005; Wang et al, 2003).

2.3 The Immunization Problems

Typical challenges in SI propagation model are as follows:

1. **Extent:** With specific subset of initially activated vertices in network and propagation model used, how many vertices are expected to be activated after a specific time/period?
2. **Targeting:** Which vertices are targeted as initiators by an adversary to result in max extent of spread? This is a hard NP to solve optimally, regardless of propagation model used [16]
3. **Blocking:** Which vertices are targeted for immunization to minimize the expected number of activated vertices [26, 27, 8] ?

3. MATERIALS AND METHODS

Dataset, Graph Model Parameters and Scope

Study samples immunization task for EVD in the Moyamba District of Southern Province in Sierra Leone. Its major towns are Moyamba, Njala, Rotifunk and Shenge – occupying a total of 2665sq mile with fourteen (14) chiefdoms. WHO report shows as of September 2014, there is no recorded/confirmed case of the disease in the district. The district has an estimated population of 260,910 (2004–2012 census) to represents our dynamic social network. Each node is an individual with a directed edge representing tie-strength and connection between two nodes. If we adopt the parameters p = 0.25, q = 0.009 and $\alpha = 6$ to generate graphs – it result in graphs with average path length and high clustering coefficient. Thus, the relationship between p, q, α and cluster coefficient as studied in Watts [32] notes α starts from 1 till it reaches 4 or more, and increase as the clustering coefficient drop so that for small values of q, high clustering coefficient is observed while clustering coefficient drops as q tends to 1.[33]

3.1 Data Gathering

The study uses the EVD spread as its innovation, so as to help tackle the immunization task. It first seeks convergence in time of final number of persons/nodes exposed to the EVD and its consequent containment as the immunization task – with these goals: (a) as participants in scheme observe WHO's safe plans, acknowledge that EVD fatality rate is high and that death is imminent, if they do not stay ahead of the outbreak, and (b) the right machinery and measures are needed to steer such new vehicle to help eradicate the EVD propagation in Sierra Leone. The study/scheme sampled individuals in the district for a 4-day programme (2-days to educate its citizen to help diffuse the info towards prevention, spread minimization and eradication of EVD). The scheme's time of introduction is its time-of-adoption and $t = 0$; and its end is $t = n$.

Chiefdoms are used as clusters as participants were further grouped using the following technique:

a. Study's time of 126days was chosen (21-days incubation period divided into 5-sections of 20% timeline each) as we seek in time, convergence of a final number of those exposed to EVD.
b. Ten (10) most influential persons in a cluster with ties in other clusters. The other/beyond is to help the nodes reach out to others outside their own cluster with the same idea. Adopt ten (10) friends to form immediate cluster – with ten (10) most influential to spread information on means of immunizing, observing and implementing WHO's safe plans. These are now involved in anti-propagation training and campaign task process.

Time-of-adoption data is based on initiation date, and is not subject to errors as some participants may state their adoption time as earlier prior to the project initiation. Such is potentially accurate and recall that errors are normally distributed.

The study does not wish to discuss the effect of the network structure on its convergence time nor the effect of clustering on the diffusion model cum process. Many studies note that innovations diffuse faster on highly clustered networks and much slowly on graphs with small degree of clustering. Our experimental model, rather aims at the expected number of final adopters as a function of the convergence task for a specified duration, the effect of tie-strengths prediction and threshold on actors' behavioral evolution and disposition towards the anti-corruption practices as the innovation.

Table 1: Dataset as Pre-Analysed on Social Network

Network Feats	EVD Immunization/Non-spread
Clusters in each community	14 clusters / chiefdoms
Targeted Number in Age range	260,910node
Time of Diffusion / Year	126days
Graph Probability Distribution	$\mu = 0.5346$ and $\delta = 0.34$
Average time of Adoption	53days
Lowest and Highest Saturation (final number of adopters)	36% and 80% at $t > 0$ and $t = t - 1$ respectively

3.2 *Rationale for ABM on Social Graph*

We use ABM to analyze relationships and behaviors in large networks for various reasons namely [21, 24]:

a. Empirically investigating preferences, disposition and other feats that make up such acts and the consequence of these acts is quite tedious and difficult.
b. Need to collect longitudinal data on entire network in time complicates empirical studies, and particularly for such complex patterns of interactions in small world-graphs.

c. People may not know why they do or took some decisions. Society representatives may not know why they are part of, behave in the particular way, and they too may be reluctant to reveal their true motivations and motives.
d. Studies suggest that emergent feats from local interactions may be biased. ABM potentials fully harnessed, models the complex properties of social systems via the analysis of data generated.
e. ABM modeled on graphs with diffusion of innovation assumes certain dispositions for a number of actors and then observes patterns as they emerge from interactions between such actors.
f. ABM involves elaborate thought experiment to learn about complex adaptive systems rather than seek to build valid representation of a real-world system. Thus, model seeks to unveil unexpected consequences of the interactions in themselves of such simple processes.

Experiment seeks to explore effects of immunization model in a number of actors as well as find the expected number of final adopters via network parameters with the actors' internal decision rules and position in the graph as they locally interact over time as exposed in the graph-based diffusion model. As such, the study is more interested in the local emergent feats emanating from large-scale effects of such interacting actors of the entire supply network.

4. EXPERIMENTAL FRAMEWORKS

Machine learning as a branch of artificial intelligence is a scientific discipline that deals with development and design of algorithms that allows machines (computers) to evolve its behaviour based on empirical data such as sensors data and databases. A learner takes advantage of data to capture its characteristics of interest of their underlying and unknown probability distribution. Such data may illustrate relationships between observed variables. Major focus on machine learning is to automatically learn to recognize complex patterns and make intelligent decisions from it [36, 30].

4.1 Fitness Function and Infection Rate

Theorem 1: With G, adjacency matrix T and infection rate $\frac{\beta}{\delta}$. $\frac{\beta}{\delta} < \frac{1}{\lambda_1(T)}$ is true, if expected time for virus to die is logarithmic. This is a function of the number of nodes in the graph against an adversary. Many interesting families of graphs holds too that $\frac{\beta}{\delta} > \frac{1}{\lambda_1(T)}$ is recovery rate – so that expected time at which virus dies out is exponential – known as Epidemic threshold.

We achieve fitness function as the immunization model aims at a solution to the network. A dynamic, non-linear model can be made linear so as to resolve it analytically. The dynamic nature of graph as social network makes them impossible to resolve analytically using non-linearity (if viewed as a multiple copies model).

Let v^t be an n-dimensional vector of states at t-steps and v_a^t is number of virus copies at node x at t-steps. Initialized at t = 0, v_a^0 is number of d copies planted in the network. At t + 1, the model evolves for (all) nodes x,y,z in the network, and for each v_a^t copies of viruses at node x, virus is propagated to node y with probability β. Virus dies when it is immunized with probability 1 – δ, and if $\Delta = \beta T + \text{diag}(1 - \delta, \ldots, 1 - \delta)$ is true, v^t is the expected state of system at time t. Then, model is completely linear if $\Delta v^t = \Delta v^{t+1}$ proven as in (Giakkoupis et al, 2010; Kempe et al, 2003). Study proposes that the objective or fitness function be achieved via Information Gain.

4.2 Bayesian Belief Model

It uses hill-climbing to search a space for optima. Once a peak is found, it restarts with another randomly chosen starting point (as such peak may not be the only one that exists). Its merit is simplicity with functions with too many maxima. Each random trial done in isolation helps immunize the nodes and overall shape of the domain is transparent to an adversary – because, as random search progresses, it continues to allocate its trials evenly over the space and evaluates as many points in the both regions found with low- and high-fitness values. Its choice is in selecting feats and attributes in graph to test is via information gain at each step while it grows the graph. The algorithm as Mitchell [19, 21]012) is thus:

DT (Examples, Target_Attribute, Attributes)
//Dataset, Attributes are feats tested by model. Target_Attribute are values to //be predicted. Return is a decision that correctly detects a given Examples.
Create a Root node of Graph
If all Examples are positive, Return single_node Graph Root with label = +
If all Examples are negative, Return single_node Graph Root with label = -
If Attribute is empty, Return single_node Graph Root, with label = most
common value of Target_Attribute in Examples
Otherwise Begin
a. A ← the attribute from attributes that best classifies Examples*
b. The decision attribute for Root ← A
c. For each possible value v_i, of A,
Add new graph branch below Root, corresponding to test A = v_i
Let Examples v_i be subset of Examples that have value v_i for A
If Examples is v_i empty
Then below this new branch, add a leaf node with label = most
common value of Target_Attribute in Examples
Else below this new branch, add the subtree
IDA(Examples v_i, Target_Attributes, Attributes – {A})
End
Return Root

The model allows each individual to be trained on partially, independently sample set of instances selected from complete training dataset. Predicted output of a classified instance is the most frequent class output of individual trees [5, 6, 19], 2002). It describes a probability distribution of a set of nodes on the graph by specifying a set of conditional independent assumptions along with a set of conditional probabilities. Thus, it allows stating conditional assumptions that simply just applies to a subset of nodes on the network by providing an intermediate and more tractable solution unlike the Naïve Bayesian Model that applies to each instance that assumptions of each graph attribute values are conditionally independent of the target value. Thus, the assumptions is that given target value of an instance, the probability of observing the interactions between nodes in the graph is the product of their probabilities from the individual attributes [7,9,10].

Note we can perform these:

a. Initialize/Select Stages: For the edge (u,v) at time t, let the corresponding state string be coded as vectors $S_u^{(t)}$ and $S_v^{(t)}$ respectively, which are interactions between nodes. Thus, we select $S_u^{(t+1)} = S_u^{(t)}$ and $S_v^{(t+1)} = S_v^{(t)}$.

b. Objective score of each new state vector is then evaluated according to the fitness function f(x). if any of them have a greater fitness value that either of their parent node, the corresponding parent nodes state vector string is replaced by its offspring for the next iteration, achieved via:

$$S_u^{(t+1)} = \underset{x \in \{S_u^{(t)}, S_u^{(t+1)}, st_1, st_2\}}{\operatorname{argmax}} f(x) \quad (1)$$

$$S_v^{(t+1)} = \underset{x \in \{S_v^{(t)}, S_v^{(t+1)}, st_1, st_2\}}{\operatorname{argmax}} f(x) \quad (2)$$

c. The immunization problem is thus defined as thus:

Problem 1 – Expected EVD Spread Targeting: Given a graph G, number of *d*-initial virus copies placed as seed nodes in each cluster and a number *k*. We observe that for *k*-nodes in G, the expected epidemic spread targeting $S'_d(G')$ in the graph is achieved as thus. As hard NP-complete task that attempts to target some nodes in G via a random strategy for influence spread, it is closely related to the sum-of-squares partition task in Aspnes et al [2] where we target a model to yield greatest propagation spread, resolved via Entropy. Entropy is a collection of nodes on G with both active (infected) and inactive (uninfected) node given by:

$$Entropy\ (E) \equiv -p_{\oplus} log_2\, p_{\oplus} - p_{\ominus}\, log_2\, p_{\ominus} \quad (3)$$

Sample n=260,910nodes, the normal/inactive/p+ = 260,000 and infected/active/seed/p- = 910. Thus Entropy is given by:

$$Entropy\,(E) \equiv -\frac{260000}{260910} log_2 \frac{260000}{260910} - \frac{910}{260910} log_2 \frac{910}{260910}$$

$$E \equiv [-(0.99) log_2 (0.99)] - [(0.01) log_2 (0.01)] = 0.004 + 0.02 = 0.024$$

This implies that innovation at time t = 1 (first 21-days of its introduction), we have that 2.4% of the population is already infected using a seed-node of 910. Thus, with an incubation period of 2 - 21days respectively, 2.4% of the people in 14 clusters will amount to 33.6% of population at t = 1. So that at t = n, the entire population is already infected if the sporadic outbreak is not prevented (see figure 1 – 4 below).[14, 15]

Problem 2: Threshold Maximization: Given G, *d* copies of nodes and infection rate of $\frac{\beta}{\delta}$, to immunize minimum number of *k* nodes in G so that $\frac{\beta}{\delta} < \frac{1}{\lambda_1(G')}$ is true. The epidemic spread given by $S'_d(G')$ in the graph is minimal. So, task attempts to immunize G with influence spread while seeking the minimal number of nodes that can be immunized.

Problem 3: Epidemic Spread Immunization: Given G, a number of *d* initial copies of viruses and *k*-nodes, if we are immunize *k*-nodes such that expected epidemic spread $S_d(G')$ in the immunized G is minimized. The role of the seed nodes planted is played by the influence-maximization model of Kempe et al [16], whose proof is omitted due to space constraint. We achieve this via Information Gain – that is the expected reduction in entropy caused by partitioning the network according to its attributes (infected and uninfected) nodes. IG is data about the target function value, given the value of another attribute A. Thus, IG of attribute (A) is:

$$Gain(E,A) \equiv Entropy(E) - \sum_{v \in Values(A)} \frac{|E_v|}{|E|} Entropy(E_v) \quad (4)$$

Values(A) is set of all possible values of Attribute A, while E_v is E and subset of attributes A with value v. Our second is expected entropy after partitioning with attribute A (sum of entropies of each subset E_v weighted by fraction of examples $\frac{E_u}{E}$ of E_v). With 33.6% infected at lowest saturation period of 20% at t = 2 and with the same number of persons being quarantined/immunized so that their treatment is extended to further cases – immunized nodes computed by Gain is thus:

$$Gain(E,A) \equiv Entropy - \sum_{v \in \{inactive, immunised\}} \frac{|E_v|}{|E|} Entropy(E_v)$$

$$Gain(E,A) \equiv 0.024 - \left(\frac{87669}{260000}\right) - \left(\frac{173248}{260000}\right) = 0.024 + \left[\left(\frac{87669}{260000}\right) + \left(\frac{173248}{260000}\right)\right] = 0.024 + [0.33 + 0.66] \equiv 1.01$$

This implies that targeting atleast 33.6% of likely infected or yet-to-be-confirmed, as seedset nodes used for immunization via the model – indicates that at *t* = 2, all nodes in the graph will be immunized as against *t* = 4. This will effectively and efficiently allows proper spread and propagation coverage.

Information Gain is further corrected via formular below:

$$Gain(E,A) \equiv Gain(E,A) \pm \left[\frac{\sum_{i=0}^{n} Gain(X_i)}{n}\right]$$

4.3 Rationale for Choice of Algorithms

Stochastic Model as inspired by evolution laws and biological population and behaviours – are rules to search domain space for optimal solution via hill-climbing method that is flexible, adaptive to changing states and suited for real-time tasks. It guarantees high convergence of optimality in multimodal task via an initialized random population and allocates increasing trials to regions found with high fitness as it finds an optimal in time. Its demerit is their inefficiency with linear systems in that if the optimal is in small region surrounded by regions of low fitness – the function becomes difficult to optimize. Thus, the adoption of the graph model is such that the *iterated hill-climbing* aims that once a peak is located, it restarts with another, randomly chosen, starting point. Its merit is simplicity and with each random trial performed in isolation, no overall picture of the domain is obtained. As the evolution progresses, it continues to allocate its trials evenly over the search space. This means that it continues to evaluate as many points in the regions found to be of low fitness as in regions found to be of high fitness.[28, 29, 31]

5. CONCLUSION

Models applied to mathematical epidemiology, focuses on analytic epidemic thresholds for varying propagation on the different families of network – seeking insight into the nature of such epidemic existence, its threshold and to unveil if such epidemic spreads or dies out [6, 5, 4]. Models serve as educational tools, to help compile knowledge about a task. They serve as modern language to communicate hypotheses, investigate parameters crucial in estimation and help us gain better insight to a problem. Their development, sensitivity and failure analysis helps reflect on functioning of natural systems.

REFERENCES

[1] Alpaydin, E., (2010). *Introduction to Machine Learning*, McGraw Hill publications, ISBN: 0070428077, New Jersey

[2] Aspnes, J., Chang, K and Yampolskiy, A., (2005). *Inoculation for victims of viruses and sum-of-squares partition problem*. In *SODA*.

[3] Axelrod R., (1997). *The complexity of cooperation*. Princeton, NJ: Princeton Univ. Press

[4] Barabasi, A.L and Albert, R., (1999). *Emergence of scaling in random networks*. Science, 286, p23.

[5] Barthelemy, M., Barrat, A., Pastor-Satorras, R and Vespignani, A. (2005). *Dynamical patterns of epidemic outbreaks in complex heterogeneous networks*. Journal of Theoretical Biology, p54.

[6] Boguna, M., Pastor-Satorras, R and Vespignani, A., (2003). *Epidemic spreading in complex networks with degree correlations*. Statistical Mechanics of Complex Networks, p36.

[7] Cohen, R., Havlin, S and Ben-Avraham, D., (2003). *Efficient immunization strategies for computer networks and populations*. Phys Rev Letters, p232.

[8] Dezso, Z and Barabasi, A.L., (2002). *Halting viruses in scale-free networks*. Phys. Rev. E 66, p67.

[9] Filiol, E., (2005). *Computer Viruses: from Theory to Applications*, New York, Springer, ISBN 10: 2287-23939-1.

[10] Ganesh, A., Massouli, L and Towsley, D., (2005). *The effect of network topology on the spread of epidemics*. In *IEEE INFOCOM*.

[11] Granovetter, M. (1978). *Threshold Models of Collective Behavior*. American J. Sociology, 83(6), p1420–1443. doi:10.1086/226707, (http://dx.doi.org/10.1086%2F226707). JSTOR 2778111

[12] Granovetter, M. (1983). *Strength of Weak Ties: A Network Theory Revisited*, Sociological Theory, 1, p201–233, doi:10.2307/202051, JSTOR 202051 (//www.jstor.org/stable/202051).

[13] Granovetter, M. (1985). *Economic action and social structure: The problem of embeddedness,* American J. Sociology, 91(3), p481–510

[14] Harrington, P., (2012). *Machine learning in action*, Manning publications, ISBN: 9781617290183, New York.

[15] Handcock, M.S and Gile, K.J., (2009). *Modeling social networks from sampled data*,* Annals of Applied Statistics, arXiv: math.PR/00000.

[16] Kempe, D., Kleinberg, J and Tardos, E., (2003). *Maximizing the spread of influence through a social network*. In *SIGKDD*.

[17] Kermack, W and McKendrick, A., (1927). *A contribution to the mathematical theory of epidemics*. Proceedings Royal Society London.

[18] Macy M and Willer, J., (2002). *From factors to actors: a computational sociology and agent based model*, Annual Review Sociology, 28, p143–166, doi: 10.1146/annurev.soc.28.110601.141117.

[19] Mitchell, T.M., (1997). *Machine Learning*, McGraw Hill publications, ISBN: 0070428077, New Jersey.

[20] Newman, M.E.J., (2003). *The structure and function of complex networks*. SIAM Reviews, 45(2), p167–256.

[21] Ojugo, A., Eboka, A., Okonta, E., Yoro, R and Aghware, F., (2012). *GA rule-based intrusion detection system*, J. of Computing and Information Systems, 3(8), p1182.

[22] Ojugo, A.A., and Yoro, R., (2013). *Computational intelligence in stochastic solution for Toroidal Queen task,* Progress in Intelligence Computing Applications, 2(1), doi: 10.4156/pica.vol2.issue1.4, p46

[23] Ojugo, A.A., Emudianughe, J., Yoro, R.E., Okonta, E.O and Eboka, A.O., (2013). *Hybrid artificial neural network gravitational search algorithm for rainfall runoff,* Progress in Intelligence Computing and Applications, 2(1), doi: 10.4156/pica.vol2.issue1.2, p22.

[24] Ojugo, A.A., Ben-Iwhiwhu, E., Kekeje, D.O., Yoro, R.E., Yerokun, M.O., (2013). *Diffusion predictor model for social graphs,* {online): www.researchgate.net/publication/261287430_decision_predictor_model_for_social_graphs, last accessed October, 2014.

[25] Pastor-Satorras, R and Vespignani, A., (2002). *Epidemics and immunization in scale-free networks*. Handbook of Graphs and Networks: From the Genome to the Internet.

[26] Singhal, P and Raul, N., (2012). *Malware detection using machine learning algorithm to assist security in Enterprise networks*, Int. J. Network Security and Applications, 4(1), p61-71.

[27] Szor, P., (2005). *The Art of Computer Virus Research and Defense*, Addison Wesley Symantec Press. ISBN-10: 0321304543, New Jersey.

[28] Toivonen, R., Kovanena, L., Kivelä, M., Onnela, J.K., Saramäkia, J and Kaskia, K., (2009). *A comparative study of social networks models: network evolution and nodal attributes models*, Social Networks, 31, p240-254, doi:10.1016/j.socnet.2009.06.004.

[29] Valente, T.W., (1996). *Social network thresholds in the diffusion of innovation*, Social Networks, 18, p69-89, SSDI 0378-8733(95)00256

[30] Wang, Y., Chakrabarti, D., Wang, C and Faloutsos, C., (2003). *Epidemic spreading in networks: An eigenvalue viewpoint*. In *SRDS*.

[31] Wasserman, S and Faust, K., (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press, pp1–27.

[32] Watts, D.J., (1999). *Networks, dynamics and the small world phenomenon*. American Journal of Sociology, 105, p234-245.

[33] World Health Organization (WHO) Fact Sheets {online}: www.who.int/csr/disease/ebola/situation-reports/en, last accessed September, 2014