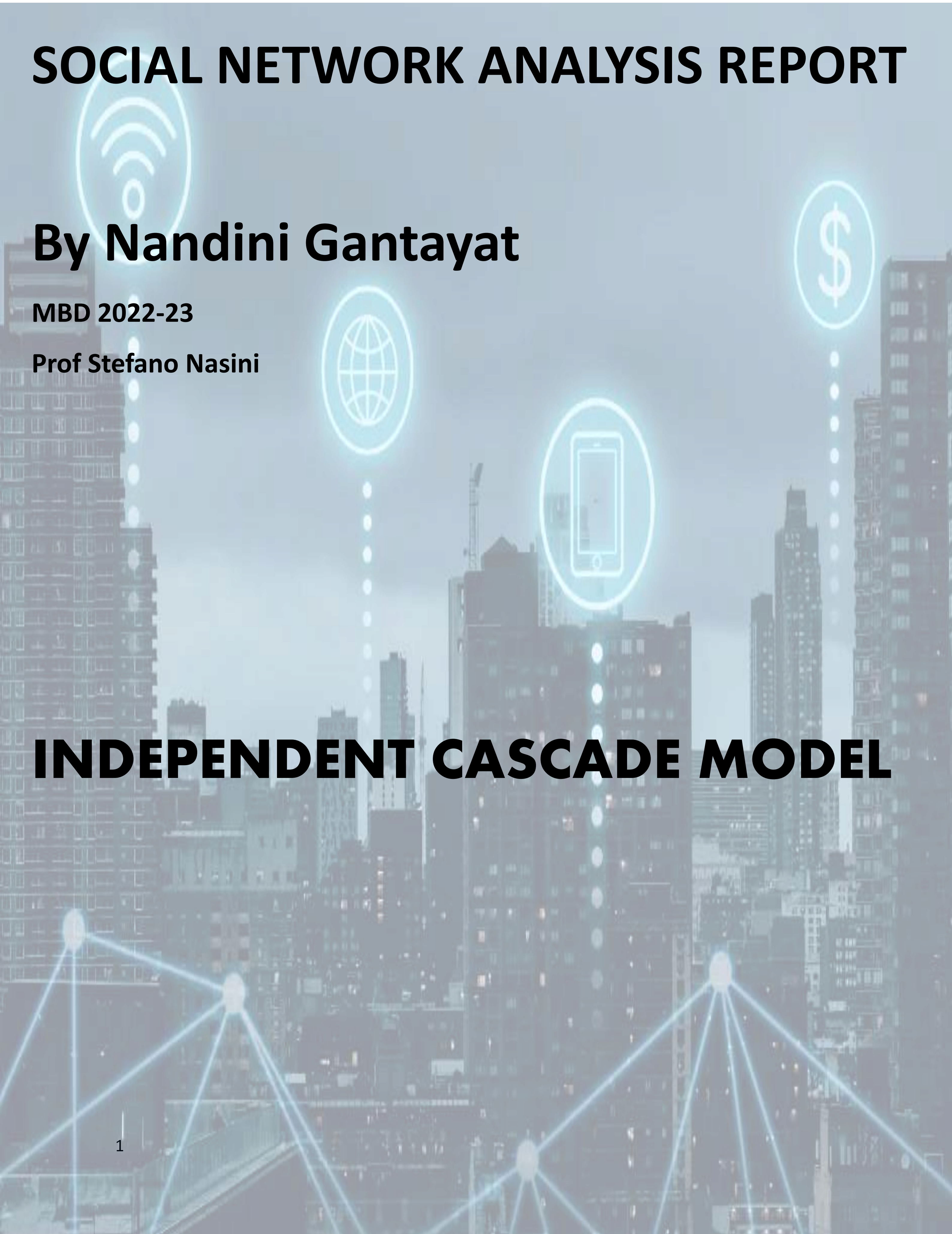


SOCIAL NETWORK ANALYSIS REPORT

By Nandini Gantayat

MBD 2022-23

Prof Stefano Nasini

INDEPENDENT CASCADE MODEL

Contents

Introduction

Cybersecurity is a pressing concern in today's interconnected world, where the rapid diffusion of information and technology poses significant risks. Understanding the dynamics of information propagation within networks is crucial for devising effective cybersecurity strategies

Stochastic models of network influence are used to study the spread of information, opinions, or behaviors in a network where randomness plays a significant role. These models take into account probabilistic factors and uncertainties in the diffusion process. Randomness is a crucial element in cybersecurity and vulnerability assessment. It adds unpredictability to systems, making them more resistant to attacks. Randomness is used in various aspects of generating cryptographic keys, simulating attack scenarios, and evaluating system resilience. It helps identify vulnerabilities that may not be apparent in specific test cases and enables organizations to develop effective mitigation strategies. Randomness is also utilized in cryptographic algorithms, access controls, and password generation to enhance security.

One of the commonly used stochastic models of network influence is the independent cascade model. It is a popular diffusion model used in network analysis. It helps to simulate the spread of information, opinions, or behaviors through a network of interconnected nodes. The model assumes that the diffusion process occurs in discrete steps or time intervals. Thus making Independent cascade models an ideal model for security threat assessments.

The Independent Cascade Model (ICM) was first introduced by Kempe, Kleinberg, and Tardos in their paper titled "Influential nodes in a diffusion model for social networks" published in the Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining in 2003.

Model description

Researchers can apply the independent cascade model to analyze network structures, study the effects of different activation probabilities and thresholds, identify influential individuals or nodes, and explore strategies to maximize or control the spread of information in a network.

Network Representation: The model starts with a network or graph representation, where nodes represent individuals or entities, and edges represent the connections or relationships between them.

Activation Probability: Each edge in the network is associated with an activation probability. This probability represents the likelihood that a node will adopt or be influenced by a particular behavior or information from its neighboring nodes.

Node Activation: The diffusion process begins with a set of "seed" nodes that are initially activated. These seed nodes could be individuals who have already adopted a behavior or received certain information.

There are three status of node:

- **Inactive** - Common nodes ;
- **Active** - seeds or nodes were activated at the last one round, these nodes can active other nodes at the next round;
- **Activated** - nodes have activated other nodes at previous round, these nodes are activated but cannot active other nodes.

Activation Threshold: Each node in the network has an activation threshold, which represents the number of activated neighbors required to activate the node. If the number of activated neighbors exceeds the activation threshold, the node becomes activated.

Cascade Propagation: At each time step, the model simulates the diffusion process. An activated node has a chance of activating its neighboring nodes based on the activation probabilities associated with the connecting edges. The activation process is independent of other cascades occurring in the network. The cascading effect in the ICM is a fundamental characteristic of diffusion models, capturing the dynamic nature of information propagation in networks. It allows for the study of how influence spreads, the identification of influential nodes, and the prediction of the overall reach of a diffusion process. The propagation of an edge will be successfully cascaded with the probability q/in_degree_of_target, where q is a hyper parameter called threshold (we always set it to 0.2, 0.4, 0.6, 0.8, or 1.0)

Cascade Termination: The cascade propagation continues until no more nodes can be activated, or a predefined stopping criterion is met. This criterion could be a maximum number of time steps, a fixed proportion of activated nodes, or other termination conditions.

Workings of the proposed model

The Independent Cascade Model is often simulated through **Monte Carlo simulation** to capture the stochastic nature of the activation process. Multiple simulations are run with different initial seed nodes to observe the average spread of influence or information in the network.

The (ICM) is often referred to as a "**diminishing return model"** because of the nature of information propagation or influence diffusion within the model, as the influence spreads from one node to its neighbors, the impact or influence tends to diminish with each subsequent step. This means that the probability of activating additional nodes decreases as the influence propagates further. The diminishing return aspect arises from the fact that each subsequent activation has a diminishing effect on the overall spread of influence in the network. Initially, the influence might have a significant impact as it reaches new nodes, but over time, the incremental influence becomes less significant as more nodes in the network become activated.

Choosing the **optimal initial seed** for a cybersecurity model in ICM, involves selecting the set of nodes that are most likely to have a significant impact on the vulnerability spread within the network

1. Prioritization based on vulnerability severity
2. Analysis of network centrality measures such as degree, betweenness, or eigenvector centrality for each node in the network.

Influence Maximization: The process of influence maximization aims to identify the initial set of nodes that, when compromised, have the maximum impact on the overall network. By simulating the spread of influence through the network, we can determine which vulnerabilities have the potential to affect the most number of nodes.

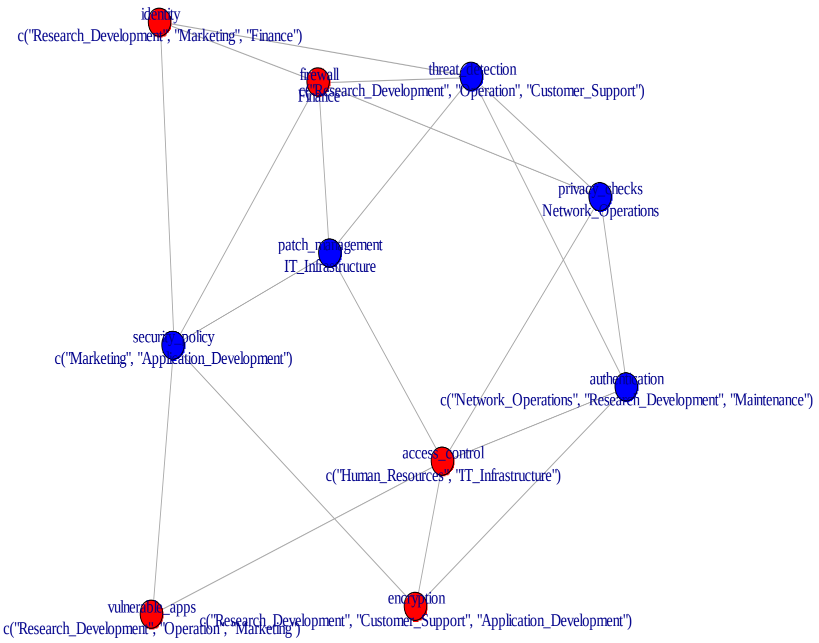
The adjacency matrix helps determine which nodes can activate others based on the model's rules. If a node representing a compromised device is activated, the adjacency matrix can identify the neighboring nodes that may be susceptible to the same attack or have a higher likelihood of being influenced.

Discrete time steps in ICM refer to the sequential progression of the diffusion process. This step-wise progression allows for the modeling of the temporal dynamics of information or threat propagation in a network.

Saturation in ICM refers to the state where a node becomes permanently activated and can no longer be influenced or deactivated by its neighbors. Once a node reaches saturation, it remains activated throughout the diffusion process. Saturation represents the point at which a node fully adopts or becomes affected by a behavior, idea, or threat in the network.

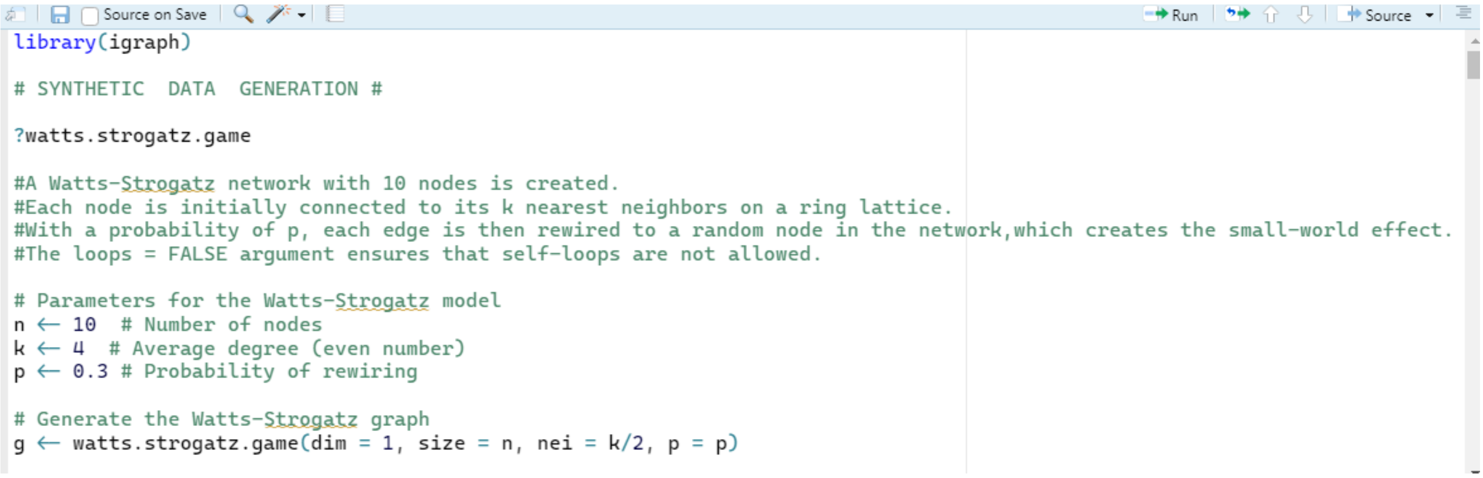
Dataset description and graphs

The generated graph g is an undirected graph. This data is generated synthetically using watts.strogatz.game from the igraph library


identity
c("Research_Development", "Marketing", "Finance")
firewall
Finance
threat_detection
c("Research_Development", "Operation", "Customer_Support")
privacy_checks
Network_Operations
patch_management
IT_Infrastructure
security_policy
c("Marketing", "Application_Development")
authentication
c("Network_Operations", "Research_Development", "Maintenance")
access_control
c("Human_Resources", "IT_Infrastructure")
vulnerable_apps
c("Research_Development", "Operation", "Marketing")
encryption
c("Research_Development", "Customer_Support", "Application_Development")

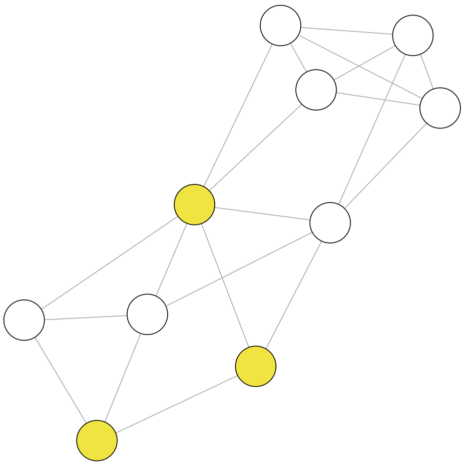
Vertex Attributes:

- V(g)$name <- node_names: Assigns the node names as the vertex attribute "name".
- V(g)$department <- node_departments: Assigns the node departments as the vertex attribute "department".

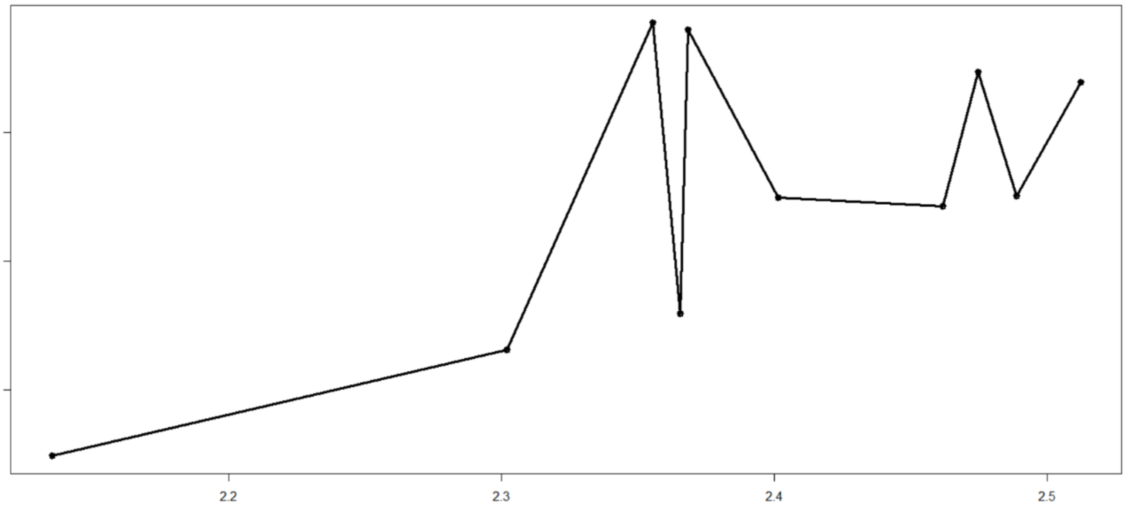


```
library(igraph)

# SYNTHETIC  DATA  GENERATION #

?watts.strogatz.game

#A Watts-Strogatz network with 10 nodes is created.
#Each node is initially connected to its k nearest neighbors on a ring lattice.
#With a probability of p, each edge is then rewired to a random node in the network,which creates the small-world effect.
#The loops = FALSE argument ensures that self-loops are not allowed.

# Parameters for the Watts-Strogatz model
n ← 10  # Number of nodes
k ← 4  # Average degree (even number)
p ← 0.3 # Probability of rewiring

# Generate the Watts-Strogatz graph
g ← watts.strogatz.game(dim = 1, size = n, nei = k/2, p = p)
```

Graphical representation of the optimal seed selection



The other graphical representation aspects include the histogram and mean and centrality measure to select the best nodes of influence



Methodology

The ICM can be described as a triple (V, E, p), where V represents the set of nodes in the network, E represents the set of edges connecting the nodes, and p is a propagation probability function that assigns a probability value to each edge (i, j) indicating the likelihood of node j being influenced by node i.

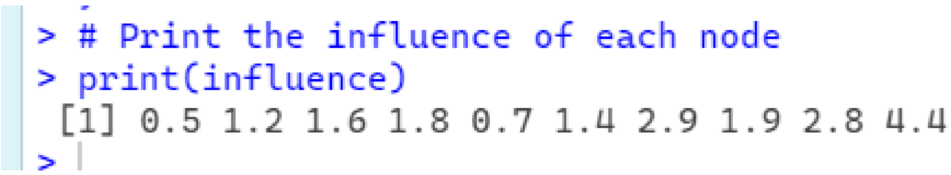
Directed Graph: Consider a directed graph G, where nodes represent different entities such as departments, systems, or network components, and edges represent the relationships or dependencies between them. The direction of the edges signifies the influence flow from one entity to another.

Source Nodes: Identify a subset of nodes, denoted as K, which represents the initial set of vulnerable nodes or potential sources of vulnerability spread. These nodes are the starting points from which the vulnerability can propagate through the network.

Probability Weight Function: Assign a probability weight function w: E → [0, 1] to each edge in the graph. This function determines the probability of vulnerability propagation from one node to its neighboring nodes. The weights can be determined based on factors such as the severity of the vulnerability, the connectivity of the nodes, or expert knowledge.

The objective of the model is to select an optimal set of source nodes S ⊆ V to maximize the expected number of influenced nodes. This can be formulated as: **maximize** |S| = K * f(S, w) where f(S, w) represents the expected number of influenced nodes when the source node set is S and the vulnerability propagation is determined by the probability weight function w.

Best Node: The "best node" refers to the initial node (vulnerability) that results in the maximum number of influenced nodes when compromised. It signifies the vulnerability with the highest potential impact on the network.



```
> # Print the influence of each node
> print(influence)
 [1] 0.5 1.2 1.6 1.8 0.7 1.4 2.9 1.9 2.8 4.4
> 
```

Further consideration

CLUSTERING, CONNECTIVITY AND HETEROGENEITY

That networks with a large average degree tend to be more robust when subjected to intentional attacks. Networks that are well-connected and have a high number of links between nodes are more resistant to deliberate attacks. On the other hand the highly clustered networks with the same degree distribution may not guarantee the same level of robustness. Clustering refers to the tendency of nodes in a network to form tightly interconnected groups or clusters. So, even if a network has a high average degree, if it is highly clustered, it may not exhibit the same level of resilience against intentional attacks.

With higher levels of connectivity, every node in the network can become a means of diffusion. This allows for a more strategic approach, focusing on those nodes that have the greatest capacity to influence others. Instead of trying to affect everyone, the emphasis shifts to targeting influential entities who can have a larger nonlinear effect on the network. This strategic targeting and leveraging of influential nodes can result in higher overall leverage and impact on the network.

Heterogeneity and clustering within the network influence the spread. Clustering creates heterogeneity, which refers to the presence of different clusters within the network. These clusters can be resistant to the spreading of a uniform phenomenon across the entire network. They may have different interpretations leading to competing variants within the network. This can create bottlenecks to the diffusion process reducing the robustness of the process and increase the capacity for exercising control over the network

NODE REMOVAL

Influence Disruption: If the removed node had a high degree of connectivity and a significant influence on the spread of infection in the network, its removal could disrupt the cascading effect and slow down the spread of the infection.

Alternate Paths of Infection: On the other hand, if the network has alternative paths of infection, removing a highly infected node may not have a substantial impact and the infection could continue to propagate.

Blocking Nodes for Rumors:

Fan et al. [3] propose algorithms to identify a minimal subset of individuals as initial protectors to minimize the number of people infected by rumors in neighbor communities.

Wang et al. [4] address the problem of minimizing the influence of rumors by discovering uninfected users. They propose a simple greedy method without theoretical analysis.

References

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5467850/#:~:text=A%20WS%20model%20%5B2%5D%20is,uniformly%20randomly%20to%20another%20node.

https://github.com/cbhua/model-independent-cascade

https://www.youtube.com/watch?v=JAfA5nkvHaM&t=1601s

https://www.youtube.com/watch?v=J1Jsh1JvztQ&t=445s

https://www.youtube.com/watch?v=rzhGTID-GD4&t=231s

https://www.youtube.com/watch?v=_-pKGNyMU-U&t=21s

Security and Communication Networks
Mathematical Models for Malware Propagation
Special Issue Editor in Chief: Angel M. Del Rey Guest Editors: Lu-Xing Yang and Vasileios A. Karyoti

Minimizing Influence of Rumors by Blockers on Social Networks: Algorithms and Analysis
Ruidong Yan, Deying Li, Weili Wu, Ding-Zhu Du and Yongcai Wang

Introduction to Stochastic Actor-Based Models for Network Dynamics
Tom A.B. Snijders, Gerhard G. van de Bunt, Christian E.G. Steglich