\documentclass[a4paper]{article}

%% Language and font encodings

\usepackage[english]{babel}

\usepackage[utf8x]{inputenc}

\usepackage[T1]{fontenc}

\usepackage{indentfirst}

\usepackage{booktabs}

\usepackage{float}

\usepackage{comment}

\usepackage{natbib}

%% Sets page size and margins

\usepackage[a4paper,top=3cm,bottom=2cm,left=3cm,right=3cm,marginparwidth=1.75cm]{geometry}

%% Useful packages

\usepackage{amsmath}

\usepackage{graphicx}

\usepackage[colorinlistoftodos]{todonotes}

\usepackage[colorlinks=true, allcolors=blue]{hyperref}

\makeatletter

\newcommand{\rmnum}[1]{\romannumeral #1}

\newcommand{\Rmnum}[1]{\expandafter\@slowromancap\romannumeral #1@}

\makeatother

%%%%%%%%%%%%%%%%%%%%%%%%%%

% Put this stuff in the header (before begin document)

%%%%%%%%%%%%%%%%%%%%%%%%%%

\usepackage{listings}

\usepackage{color} %red, green, blue, yellow, cyan, magenta, black, white

\definecolor{mygreen}{RGB}{28,172,0} % color values Red, Green, Blue

\definecolor{mylilas}{RGB}{170,55,241}

\definecolor{mygray}{gray}{0.95}

\lstset{language=R,%

%basicstyle=\color{red},

breaklines=true,%

morekeywords={matlab2tikz},

keywordstyle=\color{blue},%

morekeywords=[2]{1}, keywordstyle=[2]{\color{black}},

identifierstyle=\color{black},%

stringstyle=\color{mylilas},

commentstyle=\color{mygreen},%

backgroundcolor = \color{mygray},

showstringspaces=false,%without this there will be a symbol in the places where there is a space

numbers=left,%

numberstyle={\tiny \color{black}},% size of the numbers

numbersep=9pt, % this defines how far the numbers are from the text

emph=[1]{for,end,break},emphstyle=[1]\color{red}, %some words to emphasise

%emph=[2]{word1,word2}, emphstyle=[2]{style},

}

\title{Small Scale Diffusion Testing}

\author{Claire Kelling, Ashton Verdery}

\begin{document}

\maketitle

\thispagestyle{empty}

\begin{abstract}

Through this study, we would like to better understand how clustering in a network affects diffusion. This will help policy-makers to determine intervention strategies in given communities, after being able to draw conclusions about network vulnerability. First, we examine isomorphic networks for small regular networks. By closely examining these smaller networks, we then are able to test our diffusion functions on larger networks and this will be able to give insight on larger network structures. For these simulations, we utilize the SI and SIR diffusion models. We took the average number of time periods until the network is either saturated or no one else is infected. For these graphs, we use the complete set of isomorphic graphs. We compare the effect of the clustering coefficient and other network characteristics on the diffusion time and impact. For the second part of the project, we explore the impact of this analysis on larger scale networks through diffusion simulations. We will also briefly explore bounds on network clustering estimators.

\newline

\textbf{Keywords:} clustering coefficient, network, isomorphic graphs, diffusion modeling

\end{abstract}

\section{Introduction}

How topological features of social and other networks can affect the realization of spreading processes such as epidemics of infectious diseases, the diffusion of information, or social influence contagion that motivate behavioral change upon those networks is a topic of great interest in a variety of fields (\citep{morris1993}; \citep{rogers2010}; \citep{centola2007}). However, the nature of complex networks means that it is difficult to vary one feature without changing others, which impedes the ability to draw meaningful conclusions about the effects of varying distinct elements of network topology on the relative speed and ultimate size of the spreading process.

For instance, levels of network clustering, the tendency for one's friends to be friends with each other, are a characteristic feature of the network topology of many human social networks, with non-trivial levels of clustering endowing networks with "small world" properties (\citep{watts1998}). \citep{newman2003} studied the performance of Susceptible Infected Recovered (SIR) disease models on networks with tunable clustering levels and found that as clustering is increased, the size of the ultimate epidemic declines but the epidemic threshold, the level of infectivity needed for the epidemic to take off (hence, the speed of epidemic realization), is decreased. However, other models show the opposite properties; for example, \citep{keeling2005} finds that increases in clustering increases epidemic thresholds.

\citep{kiss2008} attempt to resolve this debate by noting that Newman's model, while varying clustering with preserved mean nodal degrees, alters levels of dispersion in the distribution of degrees such that increases in clustering lead to more degree distribution dispersion, which in turn affects the epidemic threshold, speed, and size of spreading processes. As they note, "[t]o study the theoretical effects of varying one network property (e.g., clustering), one would ideally like to generate multiple networks with all properties identical, except the property of interest. This is easier to say than do, as in practice different network properties may constrain each other, or not be independent" (\citep{kiss2008}, page 1). Of course, the examples they give regarding discrepancies between Newman's and Keeling's results illustrate this point: while Keeling's model preserves degree distributions and thus changes the conclusions about topological effects on spreading processes drawn from Newman's model, it does not constrain other macro-structural network features such as network diameter, the length of the longest shortest path.

In an effort to take a different look at this debate, in this paper we consider all isomorphic permutations of small, connected, regular random graphs. We use the complete set of all isomorphic networks with six, eight, and ten nodes (\citep{meringer1999}). According to \citep{harary1953}, "[t]wo points P and Q of a graph are called adjacent if the line PQ is one of the lines of G. Two graphs G and G’, each of n points, are called \textbf{isomorphic} if there exists a one-to-one correspondence between the points of G and those of G' which preserves adjacency." If two graphs are not isomorphic they are called \textbf{different} \citep{harary1953}. A \textbf{k-regular graph} is a graph in which each node has exactly degree k \citep{meringer1999}.

For the purposes of our work, first we look at how topological features covary with one another and then we look at diffusion processes on these networks. The focal variables we are particularly interested in are the number of triangles (clustering coefficient), the number of degrees, the cut set, and the diameter of the network.

We limit our analyses to the isomorphic regular graphs for six, eight and ten node regular networks, all with degree 3 . We believe that this will be able to give insight on larger network structures. By examining closely these smaller networks, we then are able to test our diffusion functions on larger networks.

\section{Data and Methods}

We will present some information on the two types of diffusion functions that we are considering as well as the network characteristics that we are interested in analyzing.

\subsection{Network Characteristics}

We are interested in the relationship between several key network characteristics. The basic characteristics we are interested in are diameter, number of nodes, cut set, and number of triads. This last characteristic, number of triads in the network, is to gain information about the clustering coefficient. This would be particularly useful as \citep{heath2011} have recently developed a method that takes "triangle and single edge degree sequences as input and generate a random graph with a target clustering coefficient." Therefore, if we are able to better understand the effect of the clustering coefficient on diffusion, we can use these algorithms to generate random graphs with target clustering coefficients and draw conclusions about the structural risk of that network. We can also use this analysis to compare two networks and determine comparative structural risk of these networks.

\subsection{Diffusion Functions}

The first diffusion function we use, the SIR model, takes a random node at first as infected, and with a probability of infection proceeds to infect connected nodes in the network \citep{WalkerDiff1}. However, in this first diffusion function, there is a probability of infection and if the neighboring node is not infected, the diffusion function stops and the neighboring node can no longer be infected from that starting node. The neighboring node can still be infected from other neighboring nodes, but not from the original node. In other words, this diffusion function incorporates a level of immunity, where a node can remain uninfected at the end of the process. The SIR Model, or \textbf{Susceptible-Infected-Recovery}, incorporates this same type of behavior, where individuals are susceptible and may be infected with a certain probability. After they are infected, they have a certain probability of recovery. In this case, I use a very simple SIR Model where the probability of recovery is 100\%, and they cannot be infected again.

In the second diffusion function, many of the diffusion function properties are similar. For example, there is still a probability of infection from node to node. However, this diffusion function proceeds until every node has been infected. Therefore, using this diffusion function, in the long run every node will be infected in this model. This diffusion function is called the SI Model, or \textbf{Susceptible-Infected}, where there is no chance for those who have been infected to recover.

We study the difference in time of diffusion in these two diffusion functions. In the first diffusion function, it may be of interest to also examine how many people are infected at the end. This is uninteresting in the second diffusion function because everyone is infected at the conclusion of the process. An example of each of these two functions can be seen below in Figures 1 and 2. In the graphs presented in these figures, there are two triangle, 8 nodes, and diameter 2.

%https://ezgif.com/split/ezgif-2-224aa2dc2f.gif

%http://www.online-convert.com/result/99953e14-5a5d-4e81-ba2e-9dd454fd01c5

\begin{figure}[ht!]

\centering

\includegraphics[scale=0.07]{f0.png}

\includegraphics[scale=0.07]{f1.png}

\includegraphics[scale=0.07]{f2.png}

\includegraphics[scale=0.07]{f3.png}

\includegraphics[scale=0.07]{f4.png}

\caption{SIR Model (DF1)}

\label{HN10}

\end{figure}

\begin{figure}[ht!]

\centering

\includegraphics[scale=0.07]{f0b.png}

\includegraphics[scale=0.07]{f1b.png}

\includegraphics[scale=0.07]{f2b.png}

\includegraphics[scale=0.07]{f3b.png}

\includegraphics[scale=0.07]{f4b.png}

\caption{SI Model (DF2)}

\label{HN11}

\end{figure}

For our six, eight and ten node networks, we simulate 10,000 diffusions using our two diffusion models. We took the average number of time periods until the network is either saturated or no one else is infected. The average number of people infected is either interesting or irrelevant, depending on the diffusion model. For this simulation, we used a probability of infection of 1/2. The first node that is infected is chosen at random for each simulation. For these graphs, we use the complete set of isomorphic graphs for each number of nodes (6, 8 and 10) for degree 3.

I briefly present the results for the six, eight, and ten node toy networks in table form. For all node levels, I present the average time until the process ends and the number of infected nodes at the end of the process for the first diffusion function and just the former for the second diffusion function. I then present a response surface to show how the average time changes under different sets of conditions and using both diffusion functions. To collect this data, I use the average over 10,000 runs of the simulation and a probability of infection of 1/2. The complete set of isomorphic networks used for this analysis can be seen in the Appendix.

\subsubsection{SIR Model}

\subsubsection{Six Node Graphs}

%\begin{comment}

\begin{table}[ht!]

\centering

\caption{Six Node Graph Simulations}

\begin{tabular}{ccccccc}

\toprule

Diameter & Triangles& Cut Set & Average Time & Average \# Infected \\

\midrule

2 & 2 & 3 & 4.11811 & 4.2895 \\

2 & 0 & 3 & 4.2396 & 4.3721 \\

\bottomrule

\end{tabular}

\end{table}

\begin{comment}

\begin{figure}[ht!]

\centering

\includegraphics[scale=0.5]{6node.png}

\caption{Six Node Simulation}

\label{HN2}

\end{figure}

\end{comment}

\subsubsection{Eight Node Graphs}

%\begin{comment}

\begin{table}[ht!]

\centering

\caption{8 Node Graph Simulations}

\begin{tabular}{ccccccc}

\toprule

Diameter & Triangles & Cut Set & Average Time & Average \# Infected \\

\midrule

3 & 0 & 3 & 4.6972 & 5.1783\\

3 & 2 & 3 &4.6309 & 5.0678 \\

3 & 4 & 2 &4.4859 & 4.47339 \\

2 & 0 & 3 &4.6533 & 5.1685 \\

2 & 1 & 3 &4.6521 & 5.2452 \\

\bottomrule

\end{tabular}

\end{table}

\begin{comment}

\begin{figure}[ht!]

\centering

\includegraphics[scale=0.3]{8node1.png}

\includegraphics[scale=0.3]{8node2.png}

\includegraphics[scale=0.3]{8node3.png}

\caption{8 Node Simulation}

\label{HN3}

\end{figure}

\end{comment}

\subsubsection{Ten Node Graphs}

\begin{comment}

\begin{table}[ht!]

\centering

\caption{Ten Node Graph Simulations}

\begin{tabular}{ccccccc}

\toprule

Diameter & Triangles & Cut Set & Average Time & Average \# Infected \\

1 & 5 & 4 & 1 & \\

2 & 4 & 4 & 2 & \\

3 & 4 & 3 & 2 & \\

4 & 4 & 2 & 3 & \\

5 & 3 & 4 & 2 & \\

6 & 3 & 3 & 3 & \\

7 & 3 & 2 & 3 & \\

8 & 3 & 2 & 3 & \\

9 & 3 & 2 & 3 & \\

10 & 3 & 2 & 2 & \\

11 & 3 & 1 & 3 & \\

12 & 3 & 1 & 3 & \\

13 & 3 & 1 & 3 & \\

14 & 3 & 0 & 3 & \\

15 & 3 & 0 & 3 & \\

16 & 3 & 0 & 3 & \\

17 & 3 & 0 & 3 & \\

18 & 3 & 0 & 3 & \\

19 & 2 & 0 & 3 & \\

\midrule

\bottomrule

\end{tabular}

\end{table}

\end{comment}

\subsubsection{SI Model}

\subsubsection{Six Node Graphs}

\begin{table}[ht!]

\centering

\caption{Six Node Graph Simulations}

\begin{tabular}{ccccccc}

\toprule

Diameter & Triangles& Cut Set & Average Time \\

\midrule

2 & 2 & 3 & \\

2 & 0 & 3 & \\

\bottomrule

\end{tabular}

\end{table}

\subsubsection{Eight Node Graphs}

\begin{table}[ht!]

\centering

\caption{8 Node Graph Simulations}

\begin{tabular}{ccccccc}

\toprule

Diameter & Triangles & Cut Set & Average Time\\

\midrule

3 & 0 & 3 & \\

3 & 2 & 3 & \\

3 & 4 & 2 & \\

2 & 0 & 3 & \\

2 & 1 & 3 & \\

\bottomrule

\end{tabular}

\end{table}

\subsubsection{Ten Node Graphs}

\section{Response Surface Analysis}

\subsection{Diffusion Function 1}

\subsection{Diffusion Function 2}

\newpage

\section{Bounds on Clustering and Diffusion}

\bibliographystyle{plainnat} % or try abbrvnat or unsrtnat

\bibliography{bibliography}

\section{Appendix}

\subsection{Six Node, Degree 3 Isomorphic Graphs}

\begin{figure}[ht!]

\centering

\includegraphics[scale=0.45]{6a.png}

\caption{6 Node Graphs}

\label{HN5}

\end{figure}

\subsection{Eight Node, Degree 3 Isomorphic Graphs}

\begin{figure}[ht!]

\centering

\includegraphics[scale=0.45]{Rplot01.png}

\caption{8 Node Graphs}

\label{HN8}

\end{figure}

\subsection{Ten Node, Degree 3 Isomorphic Graphs}

\begin{figure}[ht!]

\centering

\includegraphics[scale=0.45]{10a.png}

\includegraphics[scale=0.45]{10b.png}

\includegraphics[scale=0.45]{10c.png}

\caption{10 Node Graphs}

\label{HN15}

\end{figure}

\end{document}