Project 5

Logan Bolton

2025-03-30

Acknowledgement: This code was created through the repurposing of code found in the lecture notes and through collaboration with ChatGPT-4o and Gemini 2.5 Pro. AI tools were very helpful for me while fixing errors and determining the correct syntax to plot graphs.

```
# knitr::opts_chunk$set(echo = TRUE)
library('igraph')
```

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```
library(poweRlaw)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:igraph':
##
##     as_data_frame, groups, union

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
txt_file <- "/Users/log/Github/Spring2025Classes/social_networks/project6/ca-GrQc.txt"

edges <- read.table(txt_file, skip = 4, header = FALSE)
colnames(edges) <- c("FromNodeId", "ToNodeId")
g <- graph_from_data_frame(edges, directed = FALSE)
g <- simplify(g, remove.multiple = TRUE, remove.loops = TRUE)

print(paste("Number of nodes:", vcount(g)))
```

```
## [1] "Number of nodes: 5242"
```

```r
print(paste("Number of edges:", ecount(g)))
```

```
## [1] "Number of edges: 14484"
```

1) Analysis of Random Graphs G(n, p)

Fix an integer n >= 10,000. For various values of the edge probability p such that pn is a constant, use computational experiments (plots and calculations) to verify the following theoretical properties of G(n, p):

a) Mean degree c

- Verify that the mean degree c = p(n - 1).
- Plot c as a function of p.

b) Degree Distribution pk

Show that pk follows a Poisson distribution pk = e^(-c) × c^k/k! by plotting the empirical degree distribution (histogram) and overlay the theoretical Poisson curve.

c) Clustering Coefficients

Verify that both the local and global clustering coefficients of G(n, p) are equal to p.

d) Giant Component Threshold

Confirm that the threshold probability for the emergence of a giant component is 1/(n-1).

e) Fraction S of Vertices in the Giant Component

Show that the fraction S of vertices in the giant component satisfies: 1 - S = e^(-cS) by comparing the empirical value of S with the theoretical prediction (using numerical methods if needed).

f) Small Components

- Verify that small components are trees.
- Show that the average size of small components is: R = 2/(2 - c + cS)

g) Fraction of Vertices in Small Components

Verify that the fraction of vertices in small components follows (e^(-sc)*(sc)^(s-1))/s!

h) Diameter

Show that the diameter of G(n, p) follows: diameter = A + ln(n)/ln(c) where A is a constant.

2) Empirical Analysis of Real-World Network

Analyze the ca-GrQc dataset from https://snap.stanford.edu/data/ca-GrQc.html in the Stanford Large Network Dataset Collection.

a) Construct a graph G based on the data set

Analyze some basic network properties of G including order, size, density, connectivity (if G is not connected, find the number of components of G and the fraction of vertices in the largest component), and clustering coefficient.

b) Generate a configuration model G*

Generate a configuration model G* that has the same degree sequence as that of G's.

c) Analyze model properties

Analyze some basic network properties of G*.

d) Compare networks

Identify similarities and differences between G and G*.