

Modelling epidemics on networks with cliques

Anlin Zhang

Mentor: Professor Laura Schaposnik

Seventh Annual PRIMES Conference

May 20, 2017

THE PLAN

- 1 An introduction to modeling epidemics.
- 2 Modeling epidemics on regular tree graphs.
- 3 Modeling epidemics on networks with cliques.



AN INTRODUCTION TO MODELING EPIDEMICS

HISTORY

- 1 1920s: Reed-Frost model investigated the number of individuals infected and immune for certain periods of time
- 2 1927: Kermack and McKendrick's SIR model with 3 states Susceptible, Infectious and Removed
- 3 2005: Keeling and Eames's contact networks

AN INTRODUCTION TO MODELING EPIDEMICS

MODELING THROUGH GRAPHS

- Each vertex denotes an individual.
- Edges denote possible paths of infection.



Figure: In this network, pink vertices represent females, blue vertices represent males, and stars represent the origins of infection.

EPIDEMICS ON REGULAR TREE GRAPHS

Definition

A **tree** is an undirected graph in which any two vertices are connected by exactly one path, and a **rooted tree graph** is a tree in which one vertex has been designated the root. A **regular tree graph** is a tree graph in which every nonterminal vertex has the same number of edges connected to it. The **height** λ is the largest distance between any node and root, and the **degree** d is the number of nodes connected to the root by an edge.

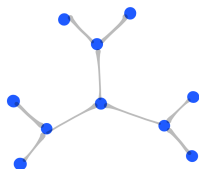


Figure: An example of a tree graph of degree $d = 3$ and height $\lambda = 2$.

EPIDEMICS ON REGULAR TREE GRAPHS

MODELING EPIDEMICS ON REGULAR TREE GRAPHS

IONTW PLATFORM

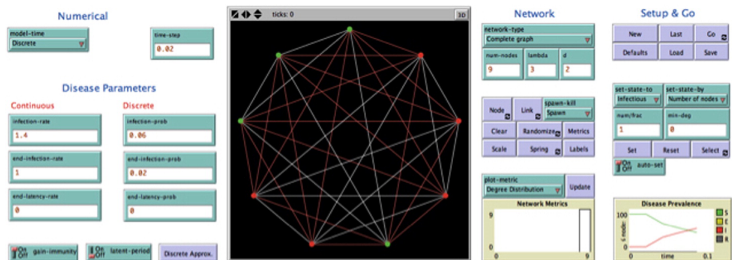


Figure: From [Seibold-Callender]

EPIDEMICS ON REGULAR TREE GRAPHS

MODELING EPIDEMICS ON REGULAR TREE GRAPHS

RELATIONS BETWEEN PARAMETERS

Proposition (Seibold-Callender)

Let N denote the number of nodes in a regular tree graph. Then

$$N = \sum_{i=0}^{\lambda} d^i.$$

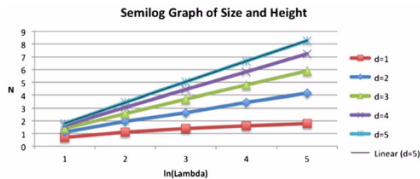


Figure: From [Seibold-Callender]

EPIDEMICS ON REGULAR TREE GRAPHS

Definition

For each node, P_{inf} is the probability of infection, while P_{rec} is the probability of recovery.

Definition

We denote by δ the number of edges in the outbreak duration. δ_λ is the expected average outbreak duration, given by $\sum_{i=1}^{\lambda+1} i \cdot P_i$, where P_i is the probability the duration spans i edges.

- Duration when $P_{rec} = 1$ and $P_{inf} = 1$
 - ▶ $\lambda + 1 \leq \delta \leq 2\lambda + 1$

EPIDEMICS ON REGULAR TREE GRAPHS

Consider $P_{inf} \neq 1$. A specific node getting infected has probability P_{inf} . Then,

- Both adjacent nodes remain uninfected with probability $(1 - P_{inf})^2$.
- One becomes infected with probability $2P_{inf} \cdot (1 - P_{inf})$.
- Both become infected with probability P_{inf}^2 .

Theorem (Seibold-Callender)

Let P_i denote the probability that the outbreak duration is i . Then

$$P_1 = (1 - P_{inf})^d$$

$$P_2 = \sum_{r=1}^d \binom{d}{r} P_{inf}^r (1 - P_{inf})^{d-r} (1 - P_{inf})^{d \cdot r}$$

$$\vdots$$

$$P_{\lambda+1} = 1 - (P_1 + P_2 + \cdots + P_\lambda).$$

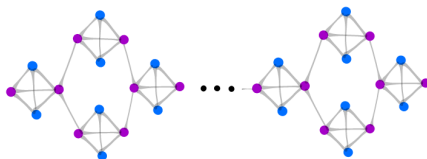
(Recall λ is the maximum distance, in edges, a node can exist from the root)

Question: How can we model outbreaks in more general types of graphs, to account for symmetries appearing, for example, from family structures?

EPIDEMICS ON REGULAR TREE GRAPHS

HOW TO EXTEND THESE RESULTS TO BETTER ENCOMPASS REAL NETWORKS?

- Irregular tree graphs
- Additional variables such as age, weight, and race
- Cliques in graphs



EPIDEMICS ON NETWORKS WITH CLIQUES

Definition

A **clique** is a subset S of a graph such that all pairs of vertices in S are connected by an edge. A clique with c vertices is called a **c -clique**. A **maximum clique** is the clique of largest size in a graph. The **clique number**, denoted as $\omega(G)$, is the number of vertices in a maximum clique of G .

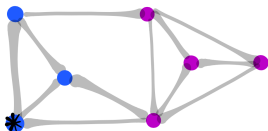


Figure: Clique number $\omega(G) = 4$.

Question: How can we bound the clique number to understand the duration of outbreaks in terms of P_{rec} and P_{inf} in this new setting?

EPIDEMICS ON NETWORKS WITH CLIQUES

INDUCED CLIQUES

Theorem (Nikiforov)

If G does not contain a complete subgraph of $k + 1$ vertices, then

$$\lambda \leq \sqrt{\frac{2(k-1) \cdot e}{k}},$$

where e denotes the number of edges.

From there, we can deduce:

Proposition

The number $\omega(G)$ is bounded by

$$\frac{1}{1 - \frac{\lambda^2}{2e}} \leq \omega(G).$$

EPIDEMICS ON NETWORKS WITH CLIQUES

REGULAR TREE GRAPHS WITH INDUCED CLIQUES

Theorem (Turan)

If G does not contain a complete subgraph of $k + 1$ vertices, then

$$e < \frac{k-1}{2k} \cdot n^2,$$

where n is the number of vertices of G .

From there, we can deduce:

Proposition

In terms of n , we bound $\omega(G)$ by

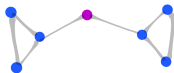
$$\frac{n}{n-\lambda} \leq \omega(G).$$

EPIDEMICS ON NETWORKS WITH CLIQUES

ROOT PLACEMENT

Question: How does the origin of infection affect outbreak duration?

- Case 1: When the root is outside all cliques



- Case 2: When the root is in a clique and connected to the "body" of G



- Case 3: When the root is a vertex inside a clique



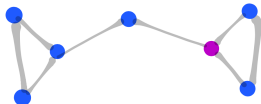
EPIDEMICS ON NETWORKS WITH CLIQUES

SAFE ZONES

Question: How can one remove a given number of links or nodes of a network so that an outbreak is contained within a subset of the graph?

Definition

The *safe zone* of a graph is defined as a subset of vertices that cannot be infected, and makes up one connected component. The number of vertices in the safe zone shall be denoted by sz .



We remove edges or vertices following the Optimal Link Removal (OLR) strategy of [Yang et al.].

SUMMARY

We have studied epidemic outbreaks modeled on graphs:

- (★) We have studied [Seibold-Callender] approach for *regular tree graphs*.
- (★) We have considered graphs with *cliques*, and bounded the clique number to understand the duration of outbreaks in terms of P_{rec} and P_{inf} in this new setting.
- (★) We have defined *Safe zones* and begun the study of their sz number.

FURTHER WORK

In what remains of the year, we will do the following:






- (★) Bound duration depending on *root placement* in networks with cliques.
- (★) Expand to include *multiple roots* of infection.
- (★) Generalize when the removal of a vertex or edge forms a *safe zone*.
- (★) Determine how to *maximize* sz .

ACKNOWLEDGEMENTS

I SINCERELY THANK...

- My mentor Professor Schaposnik for her invaluable suggestions and guidance.
- Dr. Tanya Khovanova for her help and advice.
- MIT PRIMES-USA program for this research opportunity.

MAIN BIBLIOGRAPHY

-  Nikiforov, V. "Some Inequalities for the Largest Eigenvalue of a Graph." *Combinatorics, Probability and Computing*, 11(2), 179-189.
-  Turán, Paul (1941). "On an extremal problem in graph theory", *Matematikai Értékelési Lapok* (in Hungarian), 48: 436-452.
-  Seibold, Claire, and Hannah L. Callender. "Modeling Epidemics on a Regular Tree Graph." *Letters in Biomathematics*, 30 May 2016.
-  Yang, Xiaofan, Pengdeng Li, Lu-Xing Yang, and Yingbo Wu. "Reducing the Spectral Radius of a Torus Network by Link Removal." *PLOS ONE*. Public Library of Science, 12 May 2016.
-  Kermack, W. O. and McKendrick, A. G. "A Contribution to the Mathematical Theory of Epidemics." *Proc. Roy. Soc. Lond. A* 115, 700-721, 1927.

Thank you for listening!
Any questions?