Graph Bootstrap Percolation

Case studies on selected graphs and algorithmic analysis

Andy Babb, Stephanie Martinez, Tim Wang
Advisor: Connor Anderson

Department of Mathematics, Northeastern University

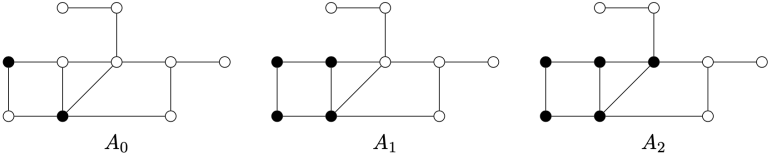
Summer REU Final Presentations – June 24, 2022

Introduction

Consider the following activation process:

- Let $G = (V, E)$ be a graph, whose vertices may be either in an "active" or "inactive" state. Fix a threshold $k \in \mathbb{N}$, and choose a seed set $A_0 \subseteq V$.
- Denoting the adjacent vertices, or neighbors of a given vertex v by $N(v)$, activations (also called infections) spread through the system as follows:

$$A_{i+1} = A_i \cup \{v \in V \ : \ |N(v) \cap A_i| \geq k\}$$



A simple example

Motivation

Bootstrap percolation can model spread in real life.

- Vertices and edges represent people and relationships, while the percolation process represents the spread of ideas, rumors, or trends.
- Which groups at minimum should be advertised to for maximal influence throughout a community?
- How can we reduce the spread of ideas or influence?

Bootstrap percolation problems

In the language of bootstrap percolation:

- Minimum contagious sets
  - What is the smallest possible seed set A_0 (and its size $|A_0|$) required to eventually infect an entire graph?
- Spread minimization
  - If we must initially infect a fixed number of vertices (i.e., choose a seed set $A_0 \subseteq V$ of fixed cardinality), which vertices should be infected to minimize spread?

Minimum contagious sets

Let $G(V, E)$ be a graph, and fix a threshold k for the percolation process.

- A contagious set is a seed set A_0 that eventually infects the entire graph (i.e., $\langle A_0 \rangle = V$).
- A contagious set is called minimum if there exists no smaller contagious set, and its size is denoted by $m(G, k)$.
- Goal: Find $m(G, k)$.
- One solution: First, compute all possible contagious sets by exhaustively finding all seed sets A_0 for which $\langle A_0 \rangle = V$. Then, $m(G, k)$ is equivalent to the cardinality of the smallest A_0 (there may be more than one).

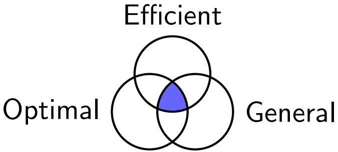
Spread minimization

Let $G(V, E)$ be a graph and fix a threshold k for the percolation process. If we must initially infect a fixed number of vertices (i.e., choose a seed set $A_0 \subseteq V$ of cardinality n), which vertices should be infected to minimize spread?

- For any seed set $A_0 \subseteq V$, we denote the set of all vertices that eventually become infected by $\langle A_0 \rangle$.
- Goal: Find a seed set A_0 of size n that minimizes $|\langle A_0 \rangle|$.
- One solution: Compute $\langle A_0 \rangle$ for all possible seed sets A_0 on the graph to find the smallest $|\langle A_0 \rangle|$ by exhaustion.

Challenges

- Spread minimization is believed to be NP-hard (though this remains unproven)
- This implies there is unlikely to be an efficient, optimal, general solution to the problem
- Hence, most approaches focus on specific cases (sacrificing generality) or finding a suboptimal solution
- Past work on the problem focuses on many different cases, supporting the suspected NP-hardness of the problem



Some interesting cases

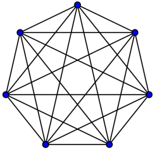
In order to better understand bootstrap percolation, we worked through percolation processes on specific kinds of graphs.

- Complete graphs
- Petersen graph
- Bipartite graphs

Complete graphs

In K_n, the complete graph of n vertices, each vertex is adjacent to every other vertex. Given a seed set A_0 and threshold k, whether any other vertices can be infected is predetermined.

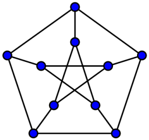
- If $|A_0| < k$, no vertices have k or more infected neighbors, so no spread occurs. Hence $|\langle A_0 \rangle| = |A_0|$.
- Otherwise, $|A_0| \geq k$, and all vertices have k or more infected neighbors. Therefore, the entire graph becomes infected in the next step, and $|\langle A_0 \rangle| = |V|$.



K_7, the complete graph of 7 vertices

Petersen graph

The Petersen graph is a peculiar graph with 10 vertices and 15 edges. It is helpful to divide its vertices into two disjoint sets: the "inner" vertices and the "outer" vertices.



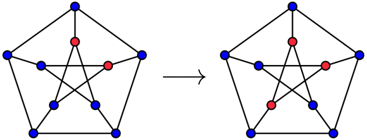
The Petersen graph

We examined outcomes for $|\langle A_0 \rangle|$ with threshold $k = 2$ for different seed sets A_0.

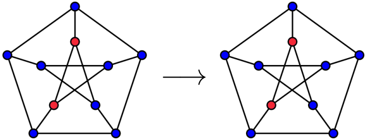
Petersen graph

Considering the "inner" vertices when $|A_0| = 2$:

- Selecting two nearby vertices results in at most one subsequent infection.



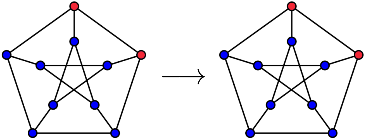
- Selecting not so nearby vertices as A_0 results in no further infections, and $|\langle A_0 \rangle| = |A_0|$.



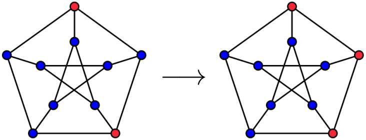
Petersen graph

Considering the "outer" vertices when $|A_0| = 2$:

- Selecting two nearby vertices results in no further infections.



- Selecting not so nearby vertices as A_0 results in at most one subsequent infection, and $|\langle A_0 \rangle| = |A_0|$.



Petersen graph

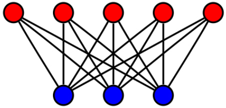
Why does this happen?

- There is an automorphism on the Petersen graph that preserves its shape, and maps the inner vertices to the outer vertices.
- Any choice of two nearby inner vertices maps to two outer vertices that are not so nearby.
- Therefore, the opposite rule applies after applying the automorphism.

Bipartite graphs

A bipartite graph $K_{m,n}$ is a graph whose vertex set V can be split into two disjoint subsets A and B, where $|A| = m$ and $|B| = n$.

- A bipartite graph is called complete if every vertex in A connects to every vertex in B.
- Percolation on the complete bipartite graph behaves similarly to the complete graph K_n.
- We explored percolation on bipartite graphs by running the percolation process on complete bipartite graphs with different combinations of m and n.



The complete bipartite graph $K_{3,5}$

Complete bipartite graphs

For $m, n \geq k$, a large enough and well distributed seed set A_0 will infect the entire graph $K_{m,n}$.

- If $|A_0| < k$, every vertex in $K_{m,n}$ has less than k infected neighbors and remains inactive.
- If $|A_0| \geq k$, A_0 can be entirely in subset A or B to infect the entire graph.
  - If A_0 is chosen to be in subset A, subset B will have $|A_0| \geq k$ infected neighbors and its vertices become activated during the first iteration of percolation.
  - During the second iteration, all vertices in A will now have $n \geq k$ infected neighbors and be activated.

General bipartite graphs

For any general bipartite graph $K_{m,n}$, $m \geq n$, if $\lceil \frac{|A_0|}{2} \rceil < k$, then we can choose a seed set such that no new vertices are activated during the percolation process.

- For a graph where $m, n \geq \lceil \frac{|A_0|}{2} \rceil$, split the seed set exactly in half. If the seed set is odd sized, let the two halves differ by 1 vertex.
- For a graph where $n < \lceil \frac{|A_0|}{2} \rceil$, activate all of subset B and as much of subset A as necessary.

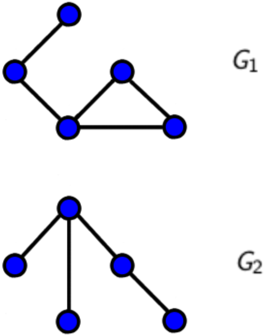
Erdős–Rényi random graphs

The Erdős–Rényi graph is a random graph represented by $G = G(n, p)$. n is the number of vertices and p is the associated probability.

- Edges of all possible combinations of vertices are added to G with probability p
- We examined only the cases in which this process creates a connected graph – any isolated vertex must be in the seed set to ever get activated
- Con: Does not accurately portray social communities

Example of Erdős–Rényi graphs

- Two possible Erdős–Rényi graphs for $G(5, 0.3)$
- All possible edges among 5 vertices are selected with probability 0.3



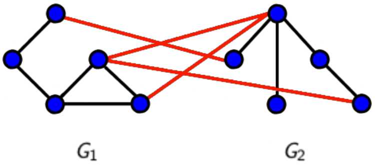
Stochastic block model

A stochastic block model (SBM) graph is the union of two Erdős–Rényi graphs G_1 and G_2. There is an additional probability q and two different communities G_1 and G_2.

- We have a graph $G = (n_1, n_2, p_1, p_2, q)$
- n_i is the number of vertices per graph
- p_i is the probability per graph for intra-connected edges
- q is the probability for inter-connected edges between communities G_1 and G_2

Example stochastic block model graph

- We union the two Erdős–Rényi graphs from before
- Edges between vertices of different communities are added with probability q



Stochastic block model

In Bootstrap percolation on the stochastic block model (Torrisi, Giovanni Luca, Michele Garetto, and Emilio Leonardi., 2022), asymptotic bounds are given for p in terms of n and k.

$$1/n_i \ll p_i \ll 1/({n_i}^{1/k})$$

Additionally, a critical number of seeds per community g_i is given in terms of p, n, and k.

$$g_i = \left(1 - \frac{1}{k}\right) \left(\frac{(k-1)!}{n_i p_i^k}\right)^{\frac{1}{k-1}}$$

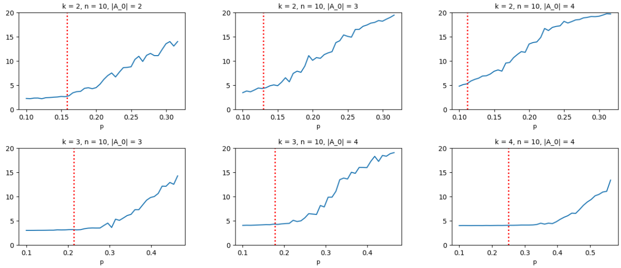
- When the the size of the seed set $|A_0| = g_i$, there is a "phase transition" in which the graph is rapidly infected

A brute-force approach

We generated computations to observe the behavior of percolation in the SBM model which includes non-asymptotic cases. Generating these computations would also allow us to see how choosing a random seed set performs for percolation. We set the following parameters,

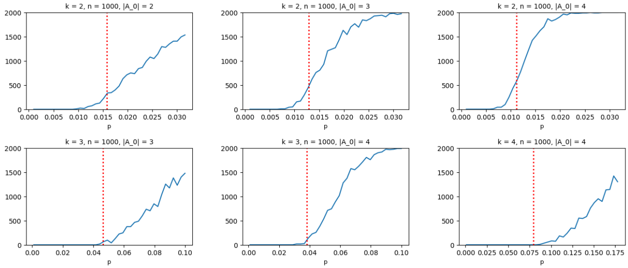
- $n_1 = n_2$ and $p_1 = p_2$
- Sizes for n of 10, 50, 100, 500, 1000 vertices
- $2 \leq |A_0| \leq 4$ and $2 \leq k \leq 4$
- k is bounded below by $|A_0|$
- q is a constant multiple of p
- 40 p-values bounded between $1/n_i$ and $1/({n_i}^{1/k})$

SBM computations, 10 vertices



Randomly chosen seed sets for $n = 10$

SBM computations, 1000 vertices



Randomly chosen seed sets for $n = 1000$

SBM Results

The red dotted line indicates when the critical number of seeds g_i is reached and equal to $|A_0|$ for some value of p. This can be derived by solving for p in $g_i = \left(1 - \frac{1}{k}\right)\left(\frac{(k-1)!}{n_i p_i^k}\right)^{\frac{1}{k-1}}$. For a randomly chosen seed set, it can be seen that,

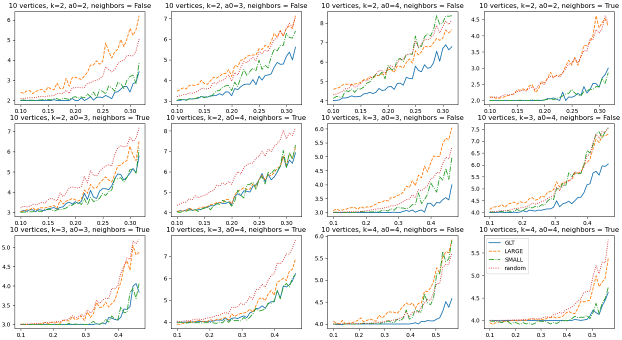
- When $|A_0| < g_i$, there is little to no spread
- When $|A_0| > g_i$, we see the phase transition in which $|\langle A_0 \rangle|$ gets larger
- So far, it looks like percolation in the SBM model performs accordingly to defined bounds!

An algorithmic approach

An attempt to choose A_0 based on degree of vertices, in a few different ways. We hoped to minimize the known number of edges between active and inactive vertices

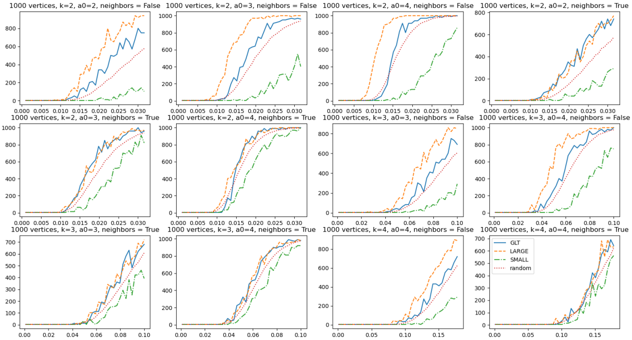
- There are two simple cases - picking the largest degree vertices, and picking the smallest degree vertices
- Can either also choose to infect the neighbors of the chosen vertices or not
- Looked into our own algorithm that chooses the vertex with "greatest degree less than remaining size of seed set"

Algorithms, 10 vertices



The algorithms on random graphs of 10 vertices

Algorithms, 1000 vertices



The algorithms on random graphs of 1000 vertices

Results

Empirically, choosing the vertices with smallest degree and not considering neighbors is the best algorithm for minimizing activation of a random graph

- Considered the same p-values as in the SBM model
- 10 vertex case is unique, potentially due to either
  - Higher variance
  - Both $|A_0|$ and k being close to the number of vertices

Next Steps

We plan to continue looking into the algorithmic approach to better understand the probability behind why certain algorithms perform as they do, and how likely algorithms are to perform within a certain threshold.

Questions

Thank you!
Any questions?