A Summary of Molloy and Reed's Theorem of Graph Percolation

Eric R. Binnendyk

University of California, San Diego
ebinnendyk@ucsd.edu

Jacobs School of Engineering
December 9, 2022

Overview

1. Introduction
   - Percolation
2. Molloy and Reed's proof
   - The theorem
   - Proof of Molloy and Reed's theorem

Introduction

Here we will summarize a result by Molloy and Reed on the connectedness of graphs generated by random models, and a related extension by Antosegui, Bonet, and Levy to satisfiability of randomly generated 2-SAT formulas.

The study of behavior of graphs generated probabilistically has gone back a long way. Erdos and Renyi gave results about the largest connected component of randomly generated graphs where each potential edge had a 1/2 probability of being chosen. Molloy and Reed's results extend this to a result about randomly generated graphs where different nodes have different expected degrees.

Percolation

Percolation is a tool used to study the connectedness of graphs.
In percolation, we start by “marking” a single node v of a graph G.
At each step, the neighbors of all marked nodes also get marked.
Eventually, we have marked an entire connected component of G.

Percolation

If the original graph is generated randomly, percolation can be regarded as a **stochastic process**, where at each step we can calculate statistics about the distribution of possible outcomes.

Percolation

In percolation theory, often a graph displays a dramatic difference in behavior after the parameters generating it pass a certain threshold, called the **percolation threshold**.
In Erdos and Renyi's model, the percolation threshold happens when $np = 1$, where p is the probability that each edge is selected and n is the number of nodes. They showed that the size of the connected components undergoes a sudden change at this point.

- When $np < 1$, the largest component is almost surely of size $O(\log n)$
- When $np > 1$, the largest component is almost surely of size $O(n)$ while no other component is larger than $O(\log n)$
- When $np = 1$, the largest component is almost surely of size $O(n^{2/3})$

The theorem

Molloy and Reed's work extends Erdos and Renyi's theorem to a situation where the graph has variable degrees of its nodes.

Theorem (Molloy and Reed, 1995)

Theorem (Connected components in graphs with variable node degrees)

If a graph G is randomly generated such that each node $v \in G$ has a particular degree $d(v)$, the percolation threshold of G is at

$$\sum_{i \geq 0} i(i-2)\lambda_i = 0$$

where λ_i is the fraction of nodes with degree i.

Theorem (Molloy and Reed, 1995)

Theorem (Connected components in graphs with variable node degrees)

If a graph G is randomly generated such that each node $v \in G$ has a particular degree $d(v)$, the percolation threshold of G is at

$$\sum_{i \geq 0} i(i-2)\lambda_i = 0$$

where λ_i is the fraction of nodes with degree i.

Theorem (Molloy and Reed, 1995)

If $\sum_{i\geq 0} i(i-2)\lambda_i < 0$, no component is larger than $O(\log n)$.
If $\sum_{i\geq 0} i(i-2)\lambda_i > 0$, the largest component takes up a positive fraction of nodes.

Proof of Molloy and Reed's theorem

In the proof, we generate a graph G dynamically.
We start with a set of values k_i that are the degree of each node $i \in G$ in our final graph G. We construct G according to a stochastic process by adding edges one at a time.

Process for generating G from the list of node degrees

- Start with all nodes disconnected.
- Choose a random node i which is connected to other nodes but isn't at degree k_i yet if possible. (Case A) Choose a node with degree 0 if there are no such nodes. (Case B)
- Choose another random node j with probability proportional to k_j minus the current degree of j.
  - In Case A1, j is connected to previously chosen nodes.
  - In Case A2, j does not yet have connections.
- Add edge (i, j) and increase degrees of i, j by 1.
- Repeat until all nodes have the required degree.

The variable X_r

In the proof, we keep track of a variable called X_r at each time step r. Let c_i be the current number of connections to node i at time r. Then:

$$X_r = \sum_{0<c_i<k_i} (k_i - c_i)$$

X_r is the number of remaining connections to add to **partially exposed nodes** — nodes which have edges but do not yet have k_i edges as required.

Three cases

What happens to the value X_r when each of the three cases is executed?

- Case B: there are no partially exposed nodes.
  - $X_r = 0$ and $X_{r+1} = k_i + k_j - 2$
- Case A1: a partially exposed node connects to another
  - $X_{r+1} = X_r - 2$
- Case A2: a partially exposed node connects to a new node
  - $X_{r+1} = X_r + k_j - 2$

Size of connected component

Cases A1 and A2 occur until $X_r = 0$, at which time case B occurs. Whenever $X_r = 0$, we have extracted a connected component of the final graph. To find the average size of a component, we find the expected time until $X_r = 0$.

When $r = o(n)$, case A2 is by far more common than case A1 as most nodes in G have not been partially exposed yet. Thus:

$$E[X_{r+1} - X_r] \approx \frac{\sum_j k_j(k_j - 2)}{\sum_j k_j} = \frac{Q(\lambda)}{E[k]}$$

$Q(\lambda) > 0$

If $Q(lambda) > 0$:

- In this case, we would expect the average value of X_r to tend upward. In fact, a basic result from random walk theory says that after $\Theta(n)$ steps, X_r is almost surely $\Theta(n)$.
- This is valid until case A1 can no longer be ignored, which happens when $r = \Theta(n)$.
- The resulting connected component almost surely has $\Theta(n)$ nodes.

$Q(\lambda) < 0$

If $Q(lambda) < 0$:

- In this case, X_r goes back to 0 fairly quickly, as it is modeled by a random walk with downward trend. Molloy and Reed show that it almost certainly happens after $O(\log(n))$ steps. Thus, if $Q(\lambda) < 0$, we produce a connected component with $O(\log(n))$ nodes.

MR proof — Conclusion

- $Q(\lambda) > 0 \Rightarrow$ giant component of size $\Theta(n)$
- $Q(\lambda) < 0 \Rightarrow$ all small components of size $O(\log(n))$

The End

Questions? Comments?