Robust Graph Analysis under Edge Uncertainty

Hanzhang Qin [1]

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology

May 17, 2017

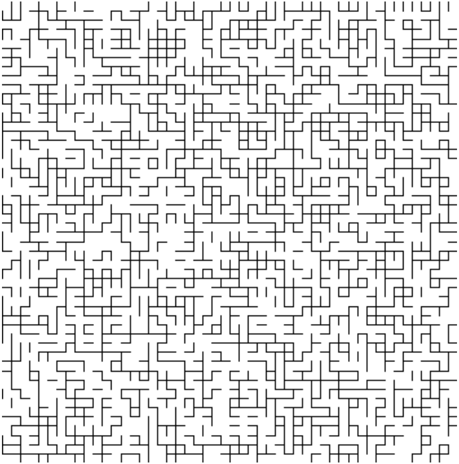
On a rainy day: the percolation theory.

- Inspired by the **percolation theory** in physics literature ([Broadbent and Hammersley, 1957], [Grimmett, 1999]).
- Suppose the rain drops are falling on a large porous stone. What is the probability that the centre of the stone is wetted?
- Classical percolation model: let $\mathbb{Z}^2$ be the (infinite) plane square lattice and let p be a number satisfying $0 \leq p \leq 1$. The edges of $\mathbb{Z}^2$ represent the inner passageways of the stone, and the parameter p is the proportion of passages which are broad enough to allow water to pass along them. So we can think all edges being "open" w.p. p, being "closed" w.p. $1 - p$.
- After deleting all the closed edges, the model reduces to a (Erdos-Renyi) random graph, with all edges present w.p. p. Then we want to analyze the structure of the graph (via the probability theory lens), e.g., the largest component size.

Any problems with this approach?

- Most analysis only applies to lattices, and hard to generalize. For example, in 1960 [Harris, 1960] conjectured there's a "threshold" of p equals $1/2$ for which there exists an infinite-sized cluster in the $\mathbb{Z}^2$ percolation model, and it took 20 more years to be proved by [Kesten, 1980].
- Different edges might behave differently in reality. In a specific graph, some edges might have high p values while others have low p values. Moreover, there might be correlation between the presence of certain edges.
- We need to estimate p in reality.

The percolation model where $p = .51$



An alternative approach: using data-driven robust optimization.

- For simplicity, we assume the presence of all edges is independent (but the result can be extended to the correlated case).
- Assume the presence of an edge (i, j) is a Bernouli random variable with parameter p_{ij}. But we do not know p_{ij}.
- Suppose we have past binary samples of (i, j): $\{I_{ij}^d, d = 1, \dots, N_{ij}\}$, then we can estimate p_{ij} as $\hat{p}_{ij} = \frac{1}{N_{ij}} \sum_{d=1}^{N_{ij}} I_{ij}^d$.
- We would like to know the graph structure within the confidence region for Pearson's χ^2 test. Namely, $(0 < \alpha < 1)$

$$\mathcal{P}_{ij}^{\chi^2} = \left\{ 0 \le p_{ij} \le 1 : \frac{(p_{ij} - \hat{p}_{ij})^2}{2p_{ij}} + \frac{(p_{ij} - \hat{p}_{ij})^2}{2(1 - p_{ij})} \le \frac{1}{2N_{ij}} \chi^2_{1,1-\alpha} \right\}.$$

- To be specific, we study edge connectivity and the largest component size of any arbitrary undirected graph under edge uncertainty.

Edge connectivity: definition

- A set of edges T is called an (a, b) edge separator if every path connecting a and b passes through at least one edge of T. Let $M(a, b)$ be the least cardinality of an (a, b) edge separator. The edge connectivity of a graph is defined as the minimum of $M(a, b)$ among all node pairs (a, b).
- Closely related to node connectivity ([Even and Tarjan, 1975]).
- Determining the edge connectivity of a graph with n nodes can be solved via a network flow model/linear optimization.

Edge connectivity: the deterministic model

- A is the adjacency matrix of the graph, i.e., $A_{ij} = 1$ if the edge (i, j) is present, and $A_{ij} = 0$ otherwise.

$$\max \; q$$

$$\text{s.t.} \; \sum_{j=1}^{n} f(j, i, u, v) = \sum_{j=1}^{n} f(i, j, u, v), \; \forall \; u < v, i \neq u, v = 1, \ldots, n,$$

$$f_s(u, v) + \sum_{j=1}^{n} f(j, u, u, v) = \sum_{j=1}^{n} f(u, j, u, v), \; \forall \; u < v = 1, \ldots, n,$$

$$\sum_{j=1}^{n} f(j, v, u, v) = f_t(u, v) + \sum_{j=1}^{n} f(v, j, u, v), \; \forall \; u < v = 1, \ldots, n,$$

$$f(i, j, u, v) \leq A_{ij}, \forall \; u < v, i, j = 1, \ldots, n,$$

$$f_s(u, v) \geq q, \forall \; u < v = 1, \ldots, n,$$

$$f_s(u, v), f(i, j, u, v), f_t(u, v) \geq 0, \forall \; u < v, i, j = 1, \ldots, n.$$

Edge connectivity: including robustness

- Consider replacing the constraint
  $f(i,j,u,v) \leq A_{ij}, \forall\ u < v, i,j = 1,\ldots,n$, by
  $f(i,j,u,v) \leq \tilde{A}_{ij}, \forall\ u < v, i,j = 1,\ldots,n,, \forall\ \tilde{A}_{ij} \sim \mathcal{P}_{ij}^{\chi^2}$.
- From [Bertsimas et al., 2013], we know the robust constraint can be further replaced by a deterministic constraint $f(i,j,u,v) \leq \delta_{ij}$ with $1-\epsilon$ probabilistic guarantee, where

$$
\begin{aligned}
\delta_{ij} = \min_{w_0,w_1,s_0,s_1,\lambda,\eta,\beta} \ & \beta + 1/\epsilon \left(\eta + \chi^2_{1,1-\alpha}/N_{ij} + 2\lambda - 2((1-\hat{p}_{ij})s_0 + \hat{p}_{ij}s_1)\right. \\
\text{s.t.}\ \ & w_0, w_1 \leq \lambda + \eta, \\
& \|[2s_0; w_0 - \eta]\|_2 \leq 2\lambda - w_0 + \eta, \|[2s_1; w_1 - \eta]\|_2 \leq 2\lambda - w_1 + \eta, \\
& - w_0, 1 - w_1 \leq \beta, \\
& s_0, s_1, w_0, w_1, \lambda \geq 0.
\end{aligned}
$$

- We can solve the resulting second-order cone optimization problem to obtain the desired result.

Purely non-probabilistic 0-1 uncertainty?

- Consider uncertainty set $\mathcal{U}^I(A, \Gamma) = \{\tilde{A} \in \{0,1\}^{n \times n} : \|A - \tilde{A}\| \leq \Gamma\}$ where $\Gamma \in \mathbb{Z}$?
- In fact less interesting.
- The behavior is quite predictable. The variation roughly equals Γ in our setting.
- Much more difficult to compute (mixed-integer optimization).

Edge connectivity: computational results

- For simplicity, we assume all $N_{ij} = 4$ and α is fixed as 0.05. We generate Erdos-Renyi random graphs (all edges are present w.p. p) as the given graphs and assume $\hat{p} = p$.
- For each parameter configuration, repeat simulating (and solving) the ER and $\mathcal{U}^I$ based model for 5 replications (note there is no need to repeat for $\mathcal{U}^{\chi^2}$ based models).

Edge connectivity: computational results

n	p	ER Edge Con.	ϵ	$\mathcal{U}^{\chi^2}$ Edge Con.	Γ	U^I Edge Con.
10	0.2	0.2	0.01	1.9	1	1
			0.05	2	3	1.4
			0.1	2.1	5	1.6
	0.5	2.4	0.01	4.7	1	2.4
			0.05	4.9	3	3.2
			0.1	5.2	5	3.6
	0.8	5.4	0.01	7.4	1	5.6
			0.05	7.6	3	6.2
			0.1	8.1	5	6.6
20	0.2	0.6	0.01	4.1	1	0.8
			0.05	4.3	3	1.6
			0.1	4.5	5	2.4
	0.5	5.4	0.01	9.9	1	6
			0.05	10.3	3	6.4
			0.1	10.9	5	7.4
	0.8	12	0.01	15.6	1	12
			0.05	16.2	3	12.8
			0.1	17.1	5	13.4
30	0.2	1.2	0.01	6.2	1	1.6
			0.05	6.5	3	2
			0.1	6.9	5	2.2
	0.5	9.6	0.01	15	1	10.4
			0.05	15.7	3	11
			0.1	16.6	5	11.6
	0.8	19.2	0.01	23.8	1	20.1
			0.05	24.8	3	20.8
			0.1	26.2	5	21.6

Largest component size: computational results

- This problem can be cast as a very similar network flow model just as we have seen. So we omit the specific formulations.

Largest component size: computational results

n	p	ER Large. Comp. Size	ϵ	$\mathcal{U}^{\chi^2}$ Large. Comp. Size	Γ	$\mathcal{U}^{I}$ Large. Comp. Size
10	0.01	1.6	0.01	2.1	1	2.6
			0.05	2.2	2	3.8
			0.1	2.2	3	4.8
	0.02	1.8	0.01	3	1	3
			0.05	3.1	2	4
			0.1	3.2	3	5
	0.05	2.2	0.01	5.7	1	4.6
			0.05	5.9	2	6
			0.1	6.2	3	6.8
20	0.01	1.8	0.01	6	1	2.4
			0.05	6.2	2	3.2
			0.1	6.5	3	4
	0.02	3.2	0.01	10.1	1	4
			0.05	10.5	2	5
			0.1	11	3	6
	0.05	7.6	0.01	20	1	8
			0.05	20	2	8.4
			0.1	20	3	9
30	0.01	2.6	0.01	12.6	1	3.2
			0.05	13.1	2	4
			0.1	13.8	3	4.6
	0.02	3.8	0.01	22.2	1	5
			0.05	23.1	2	5.4
			0.1	24.3	3	6.4
	0.05	18.6	0.01	30	1	19.8
			0.05	30	2	20.2
			0.1	30	3	21

Endnotes

- Both robust models give more conservative estimates.
- The $\mathcal{U}^{\chi^2}$ based robust model mimics the stochastic Erdos-Renyi model pretty well.
- We can easily generalize this approach to many other graph characteristics, e.g., the diameter, total number of degrees.
- The network flow problems might have better formulations.

Bertsimas, D., Gupta, V., and Kallus, N. (2013).
Data-driven robust optimization.
arXiv preprint arXiv:1401.0212.

Broadbent, S. R. and Hammersley, J. M. (1957).
Percolation processes: I. crystals and mazes.
In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 53, pages 629–641. Cambridge University Press.

Even, S. and Tarjan, R. E. (1975).
Network flow and testing graph connectivity.
SIAM journal on computing, 4(4):507–518.

Grimmett, G. (1999).
What is percolation?
In *Percolation*, pages 1–31. Springer.

Harris, T. E. (1960).
A lower bound for the critical probability in a certain percolation process.

In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 56, pages 13–20. Cambridge University Press.

Kesten, H. (1980).
The critical probability of bond percolation on the square lattice equals $1/2$.
Communications in mathematical physics, 74(1):41–59.