Limitations of Chung Lu Random Graph Generation

Christopher Brissette [1] George Slota [1]

[1]Renssealer Polytechnic Institute, Troy NY

November 11, 2021

The Chung-Lu Random Graph Model

- We focus on a particularly simple type of stochastic block model.
- The Chung-Lu Random graph model[1] can be thought of as a generalization of the Erdös Reyni random graph model that allows for degree distributions to be matched in expectation.
- It is a SBM with communities consisting of nodes with the same expected degree w_i.

$$p_{ij} = \frac{w_i w_j}{\sum_{i \in V} w_i}$$

[1] Chung and Lu 2002.

The Chung-Lu Random Graph Model

- Note that generating a graph with Chung-Lu may be done embarrassingly parallel.
- This is in opposition to the standard configuration model which matches a degree distribution explicitly[2], however has limited room for parallelism.

Partially due to this, as well as the simplicity of implementing Chung-Lu, it is a popular subroutine for more complex graph generation algorithms.

[2] Fosdick et al. 2018.

The Chung-Lu Random Graph Model

For instance BTER[3] which matches a degree distribution and clustering coefficient distribution in expectation. This algorithm takes place in three steps.

1. Partition nodes into "affinity blocks".
2. Connect block internals according to calculated ER probabilities.
3. Connect between blocks using Chung-Lu probabilities.

[3] Kolda et al. 2014.

Chung-Lu shortcomings

- The authors of BTER[4] note that there is an issue with generating low-degree nodes and implement additional functionality to deal with it.
- In fact, many authors[5] point out that Chung-Lu has difficulty accurately recreating certain degree distributions.

[4] Kolda et al. 2014.

[5] Winlaw, DeSterck, and Sanders 2015; Britton, Deijfen, and Martin-Löf 2006; Hofstad 2013; Pfeiffer III et al. 2012; Durak et al. 2013.

Limits of Chung-Lu

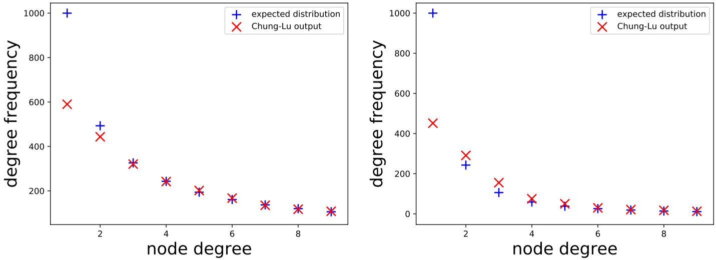

Figure: Output distributions of Chung-Lu given two power law degree distributions with exponents $\beta = 1.0$ and $\beta = 2.0$. Note the inaccuracies on the low-degree nodes.

The "Shifted" Chung-Lu solution

Question: How might we better match degree distributions while retaining the simplicity of Chung-Lu and its capacity for parallelism?

- Other work[6] has worked on improving the accuracy of Chung-Lu in the case of directed graph generation.
- They manage error in degree one nodes with the use of a "blowup factor".
- This amounts to increasing the number of nodes in the degree-one block without increasing their probability of selection.

[6] Durak et al. 2013.

The "Shifted" Chung-Lu solution

Question: How might we better match degree distributions while retaining the simplicity of Chung-Lu and its capacity for parallelism?

Our theoretical contribution:

1. Create a linear model (matrix) representing Chung-Lu.
2. Invert that linear model.
3. Use this inverse to determine the appropriate input distribution for Chung-Lu.
4. Run Chung-Lu on this input.

The "Shifted" Chung-Lu solution

How might we make a linear representation of Chung-Lu? We note a few properties.

- We expect each degree block to be binomially distributed.
- For sufficiently large degree blocks we can approximate this as a Poisson distribution with mean w_i. (This is also noted in[7])

[7] Durak et al. 2013; Karrer and Newman 2011.

The "Shifted" Chung-Lu solution

Within a degree block B_w of mean w, the number of nodes we expect to have of any given degree k is then described by

$$poiss(w, k)|B_w|$$

This gives us a natural way to turn our problem into solving a linear system.

- Form a matrix $\mathbf{P}$ with Poisson distributions having means $1, 2, \cdots, w_{max}$ as the columns.
- Then degree distributions can be input as column vectors $\mathbf{x}$ to be input into the model.

The "Shifted" Chung-Lu solution

There are a couple things to note.

1. We need our matrix to be square, so we truncate it to some fixed dimension $\mathbb{R}^{w_{max}}$.
2. We need to ensure that this matrix is in fact invertable.
3. We need to classify when the inverse of $\mathbf{P}$ maps into negative dimensions in $\mathbb{R}^{w_{max}}$.

- Note that $\mathbf{P} = \mathbf{AVB}$ is a factorization of $\mathbf{P}$ where $\mathbf{A}$ and $\mathbf{B}$ are diagonal real matrices and $\mathbf{V}$ is a Vandermonde matrix on integer nodes.
- Invertability is guaranteed as a consequence of work with Vandermonde matrices[8].

[8] Eisinberg, Franzé, and Pugliese 1998.

The "Shifted" Chung-Lu solution

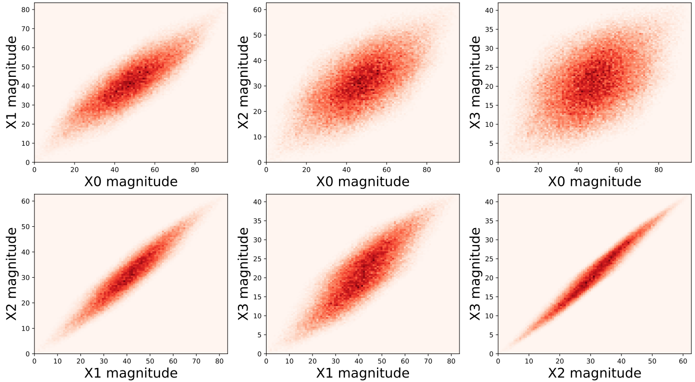

Figure: An example of "valid" distributions calculated by applying **P** on random vectors in the positive region of $\mathbb{R}^4$. The hot-red regions show 2D projections of this valid region given these random input vectors.

The "Shifted" Chung-Lu solution

While figuring out if a given distribution is valid for inversion or not will generally require explicit computation, we can classify "extreme" distributions beyond which no valid distributions can exist.

- Note that the diagonal elements of **P** are maximal within their row and column.
- This means that $(\mathbf{Px})_i \leq \mathbf{P}_{ii} \|x\|_1$.
- This tells us that ER graphs with the correct probabilities maximize the number of nodes with a given degree.
- More interestingly, this puts hard limits on the degree distributions we can expect to generate with Chung-Lu. (For instance we should never expect more than $\approx 17\%$ of nodes to have degree 5)

The "Shifted" Chung-Lu solution

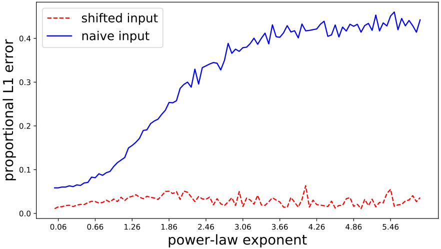

Figure: proportional $L1$ distribution errors for naïve Chung-Lu inputs versus shifted Chung-Lu inputs for power-law distributions with $w_{max} = 40$.

A more detailed look at this work is currently housed on the arXiv[9] and will be published in the conference proceedings.

[9] Brissette and Slota 2021.