Hierarchical Latent Space Models

Alex Martin Loewi
aloewi@cmu.edu

Francisco Ralston Fonseca
fralston@andrew.cmu.edu

Octavio Mesner
omesner@cmu.edu

Abstract

abstract...

1 Introduction

Testing citations [1], [2, 3], [4, 5, 6].

Social network analysis (SNA) is a very strange field, in a number of ways. it is inherently interdisciplinary, even its name indicating that it emphasizes both sophisticated statistical machinery, and also much more qualitative elements. While this is an exciting intersection, spanning both of these such disparate approaches is exceptionally difficult. As a result many important problems in SNA are under-studied, simply because of the small number of people who think about the problems it considers, in their full depth.

One of these problems is that of modeling multigraphs. Multigraphs, also called multiview networks, can be thought of as either a single graph with multiple edge types, or as multiple networks over the same set of nodes. Data of this kind is not uncommon, at either large or small scales. Twitter has Following relationships, as well as Retweets, and Likes, each of which can be thought of as simply one layer within the full relationship between two nodes. In small scales, surveys for networks very rarely ask about a single type of relationship, and many of the most famous network data sets have multiple relationships (such as Trust, Friendship, and Respect, in Samson's monk data). Despite the existence of data of this form for decades, the strange nature and inherent difficulty of this data has led to an extremely small number of ways to model it.

In designing a model however, the complexity of the data requires a perspective on what aspect of the network is most important. The model considered here was inspired initially by a very practical problem. A particular data set had five layers, and the goal was to compare visualizations of the five separate layers, to gain an intuition for similarities in the networks (visualization has long been a cornerstone of SNA, and graph layout algorithms is a substantial field unto itself). However, even if it has the same set of nodes, a graph with different sets of edges may be laid out completely differently – incomparably – from a graph with a different set of edges. The other approach, to immediately solve this problem of incomparable node placement, is to fix all of the positions a priori. However it has also long been known that doing so is the least informative visualization possible, as the very point of layout algorithms is to drive placement as a function of connectivity. This inspired the thought that perhaps a model could trade off these two extremes – allowing just enough flexibility that the layout of a single layer could be both informative, and also compared meaningfully with the other layers in a graph.

2 Problem Statement

2.1 The Likelihood function

The Binomial likelihood equation of the HLSM is:

$$L = \prod_k \prod_i \prod_{j<i} \sigma(\eta_{ijk})^{y_{ijk}} (1 - \sigma(\eta_{ijk}))^{1-y_{ijk}}$$

where,

$$\eta_{ijk} = \alpha_k - \|z_{ik} - z_{jk}\|^2 = \alpha_k - \|b_i + \epsilon_{ik} - b_j - \epsilon_{jk}\|^2$$

$$z_{ik} = b_i + \epsilon_{ik}$$

$$\sigma(\eta_{ijk}) = \frac{1}{1 + \exp(-\eta_{ijk})}$$

Then, the log likelihood is

$$
\begin{aligned}
\ell &= \sum_k \sum_i \sum_{j<i} y_{ijk} \ln \sigma(\eta_{ijk}) + (1 - y_{ijk}) \ln(1 - \sigma(\eta_{ijk})) \\
&= \sum_k \sum_i \sum_{j<i} -y_{ijk} \ln(1 + \exp(-\eta_{ijk})) + (1 - y_{ijk}) \ln \left(\frac{\exp(-\eta_{ijk})}{1 + \exp(-\eta_{ijk})} \right) \\
&= \sum_k \sum_i \sum_{j<i} -y_{ijk} \ln(1 + \exp(-\eta_{ijk})) + (1 - y_{ijk})[-\eta_{ijk} - \ln(1 + \exp(-\eta_{ijk}))] \\
&= \sum_k \sum_i \sum_{j<i} -y_{ijk} \ln(1 + \exp(-\eta_{ijk})) - \eta_{ijk} - \ln(1 + \exp(-\eta_{ijk})) + y_{ijk}\eta_{ijk} + y_{ijk} \ln(1 + \exp(-\eta_{ijk}))] \\
&= \sum_k \sum_i \sum_{j<i} (y_{ijk} - 1)\eta_{ijk} - \ln(1 + \exp(-\eta_{ijk}))
\end{aligned}
$$

2.2 Optimization Problem

Finally, our complete objective function is to minimize the sum between the negative log likelihood and a lasso penalty on the deviations ϵ_{ik}.

$$\min_{\epsilon,b} \sum_k \sum_i \sum_{j<i} [(1 - y_{ijk})\eta_{ijk} + \ln(1 + \exp(\eta_{ijk}))] + \lambda \sum_k \sum_i \|\epsilon_{i,k}\|$$

where $\|.\|$ represents the L2 norm. Since this is not differentiable, we could use proximal gradient descent. If the gradients for the ML function derived above are correct, we can directly use the code we implemented in HW3 (problem 4) to implement the proximal operator for this problem.

2.3 Derivation of the gradients

We define our decision variables as: $\{b_i\} \; \forall \, i$ and $\{\epsilon_{ik}\} \; \forall \, i, k$.

Then, the gradients of the η_{ijk} are:

$$\frac{\partial \eta_{ijk}}{\partial b_i} = -2(b_i + \epsilon_{ik} - b_j - \epsilon_{jk})$$

$$\frac{\partial \eta_{ijk}}{\partial b_j} = 2(b_i + \epsilon_{ik} - b_j - \epsilon_{jk})$$

$$\frac{\partial \eta_{ijk}}{\partial \epsilon_{ik}} = -2(b_i + \epsilon_{ik} - b_j - \epsilon_{jk})$$

$$\frac{\partial \eta_{ijk}}{\partial \epsilon_{jk}} = 2(b_i + \epsilon_{ik} - b_j - \epsilon_{jk})$$

$$\frac{\partial \ell}{\partial b_m} = \sum_k \sum_i \sum_{j<i} (y_{ijk} - 1) \frac{\partial \eta_{ijk}}{\partial b_m} - \frac{\partial \ln(1 + \exp(-\eta_{ijk}))}{\partial b_m}$$

$$= \sum_k \left[\sum_{j<m} (y_{mjk} - 1) \frac{\partial \eta_{mjk}}{\partial b_m} - \frac{\partial \ln(1 + \exp(-\eta_{mjk}))}{\partial b_m} + \sum_{i>m} (y_{imk} - 1) \frac{\partial \eta_{imk}}{\partial b_m} - \frac{\partial \ln(1 + \exp(-\eta_{imk}))}{\partial b_m} \right]$$

$$= \sum_k \left[-2 \sum_{j<m} \left[(y_{mjk} - 1)(b_m + \epsilon_{mk} - b_j + \epsilon_{jk}) + \frac{\exp(-\eta_{mjk})}{1 + \exp(-\eta_{mjk})}(b_m + \epsilon_{mk} - b_j - \epsilon_{jk}) \right] \right.$$

$$\left. + 2 \sum_{i>m} \left[(y_{imk} - 1)(b_i + \epsilon_{ik} - b_m - \epsilon_{mk}) + \frac{\exp(-\eta_{ijk})}{1 + \exp(-\eta_{mjk})}(b_i + \epsilon_{ik} - b_m - \epsilon_{mk}) \right] \right]$$

$$= 2 \sum_k \left[- \sum_{j<m} \left(y_{mjk} - 1 + \frac{\exp(-\eta_{mjk})}{1 + \exp(-\eta_{mjk})} \right) (z_{mk} - z_{jk}) \right.$$

$$\left. + \sum_{i>m} \left(y_{imk} - 1 + \frac{\exp(-\eta_{ijk})}{1 + \exp(-\eta_{mjk})} \right) (z_{ik} - z_{mk}) \right]$$

$$\frac{\partial \ell}{\partial \epsilon_{mk}} = 2 \left[- \sum_{j<m} \left(y_{mjk} - 1 + \frac{\exp(-\eta_{mjk})}{1 + \exp(-\eta_{mjk})} \right) (z_{mk} - z_{jk}) \right.$$

$$\left. + \sum_{i>m} \left(y_{imk} - 1 + \frac{\exp(-\eta_{ijk})}{1 + \exp(-\eta_{mjk})} \right) (z_{ik} - z_{mk}) \right]$$

2.4 The proximal operator

The proximal operator for the penalty term using the L2 norm[1] is:

$$\text{prox}_{\|\cdot\|,t}(\epsilon_{i,k}) = \begin{cases} \frac{\|\epsilon_{i,k}\| - \lambda t}{\|\epsilon_{i,k}\|} \epsilon_{i,k}, & \|\epsilon_{i,k}\| \geq \lambda t \\ 0, & \|\epsilon_{i,k}\| < \lambda t \end{cases}$$

$$\text{prox}_{\|\cdot\|,t}(b_i) = b_i$$

Therefore, let β be the vector with all parameters ϵ and b. The Proximal Gradient Descent method in this case is:

$$\beta^+ = \text{prox}_{\|\cdot\|,t}(\beta - t \nabla \ell(\beta))$$

3 Results

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

---

[1]According to the results on Homework 3

4 Discussion

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

5 Conclusion

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

References

[1] Paul R. Rosenbaum. Optimal matching for observational studies. *Journal of the American Statistical Association*, 84(408):1024–1032, 1989. ISSN 01621459. URL http://www.jstor.org/stable/2290079.

[2] Joseph D. Y. Kang and Joseph L. Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.*, 22(4): 523–539, 11 2007. doi: 10.1214/07-STS227. URL https://doi.org/10.1214/07-STS227.

[3] Kosuke Imai and Marc Ratkovic. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):243–263, 2014. ISSN 1467-9868. doi: 10.1111/rssb.12027. URL http://dx.doi.org/10.1111/rssb.12027.

[4] Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002. doi: 10.1198/016214502388618906. URL https://doi.org/10.1198/016214502388618906.

[5] Isabella Gollini and Thomas Brendan Murphy. Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265, 2016. doi: 10.1080/10618600.2014.978006. URL https://doi.org/10.1080/10618600.2014.978006.

[6] Michael Salter-Townshend and Tyler H. McCormick. Latent space models for multiview network data. *Ann. Appl. Stat.*, 11(3):1217–1244, 09 2017. doi: 10.1214/16-AOAS955. URL https://doi.org/10.1214/16-AOAS955.