

EM and Three Dimensional Block Modeling in Cognitive Social Structures

Alexander Loewi

April 5, 2015

Evidence continues to accumulate that two people in the same social network can have radically different perceptions of that network. They disagree on who is more central, who is friends with whom, and even whether they have a relationship themselves. Whether this comes as a deep surprise or sounds like the most obvious thing in the world, this insight has generated an enormous amount of interest over the last three decades.

While there is no question that the findings are qualitatively compelling, there is also no question that the data that represents them is quantitatively difficult. While many techniques have been developed for operating on single social networks, there are far fewer techniques for dealing with several networks at once, and in particular the multiple networks that represent different perceptions. In addition, those few techniques that have been developed tend to suffer from two serious problems. First, they tend not to be well integrated into familiar statistical frameworks, meaning their theoretical properties are often still in question. Second, they often operate by forcing the multiple networks into a single network, potentially losing all of the structure that makes the differences interesting.

This paper takes two extremely well established statistical methods – Expectation Maximization and Stochastic Block Models – and uses them to perform a number of important tasks on the totality of Cognitive Social Structure (network perception) data. Broadly, these tasks fall into the categories of edge inference, and group discovery, although both of these characterizations have some nuance to them. The methods developed here allow for a principled approach to missing data – an almost universal problem, when trying to collect network perceptions – as well as a way to find more meaning in complete data.

To demonstrate their utility, these methods are tested on several famous datasets, in order to answer three questions – what is the edge inference accuracy with different amounts of randomly sampled data, what is the group discovery accuracy with different amounts of randomly sampled data, and finally, if data can be collected based on the existing data instead of randomly, can this be used to produce faster convergence in edge or group inference?

1 Modeling Cognitive Social Structures

As we go about modeling this data, there are two primary concerns – what features of the data are we trying to describe, and what network models exist that are amenable to these features?

1.1 The Correlates of Error in Cognitive Social Structures

Network perceptions are interesting precisely because of their errors, and so in order to understand them, it is vital to model this error. The most common statistically specified error process is a completely random one, which in this case would mean that every person had an equal and fixed chance of misperceiving an edge. Both intuitively and empirically however, this process fails to capture several important features of the observed errors.

Perhaps most predictably, people are more accurate when it comes to perceiving relationships that are “close” to them socially. In other words, you are more likely to accurately perceive one of your own relationships, or one of your friend’s, than of two people you never talk to. Another consistent source of error is that people are not equally likely to make false positive and false negative errors. Overall, there is a significantly greater tendency to fail to observe an existing relationship (commit a false negative) than there is to imagine one that doesn’t exist (commit a false positive). The last consistent finding is that individuals have very different error rates from one another – beyond the effects of distance and omission/commission biases, people appear to have a personal tendency to accurately perceive relationships or not, and this tendency might be very different from someone else’s. Again, this is not surprising – everyone knows someone who just has difficulty reading social situations – but the degree to which this can be observed in the data is important to emphasize.

1.2 Blockmodels of Cognitive Social Structures

With the desire to use models that are well established and understood, there are just a few broad classes of network models to choose from. These are Exponential Random Graph Models (ERGMs), Stochastic Block Models (SBMs) and Latent Space Models (LSMs) – the latter two of which have also been further generalized into the single category of Conditional Independence Models (CID models).

1.3 Three Dimensional Blockmodels

2 Edge Inference Via Stochastic Blockmodeling

$$p(\text{edge} \mid \text{observation}) = \frac{p(e, o)}{p(o)} = \frac{1}{p(o)} \quad (1)$$

2.1 Expectation Maximization within SBMs

Algorithm 1 The Structure of the EM Procedure

```

 $\hat{G}_0 \leftarrow \text{INIT}(X)$ 
while True do
     $B_n \leftarrow \text{BLOCKMODEL}(\hat{G}_n)$  ▷ M-step
     $\hat{G}_{n+1} \leftarrow \text{POSTERIOR}(X, B_n)$  ▷ E-step
    if  $\|\hat{G}_n - \hat{G}_{n+1}\| < \epsilon$  then
        BREAK
    end if
end while
RETURN( $\hat{G}$ )

```

3 Three Dimensional Stochastic Block Modeling

$$\sum_{rs} m_{rs} \log \frac{m_{rs}}{n_r n_s}$$

3.1 Extending Degree Correction to Three Dimensions

The primary difficulty in extending the degree corrected model to three dimensions is a conceptual one. An edge is an intuitive object, as pointing at things is universal and ubiquitous. In three dimensions however, an edge has not only a head and a tail, but also a midpoint, which successfully confounds that valuable intuition. The productive way forward is to abandon it, and think of an edge only as the mathematical abstraction of an ordered tuple – in this case, a 3-tuple.

$$\sum_{rst} m_{rst} \log \frac{m_{rst}}{\kappa_{r,1} \kappa_{s,2} \kappa_{t,3}}$$

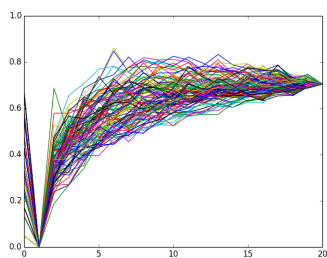
4 The Data

4.1 Global Error Rates

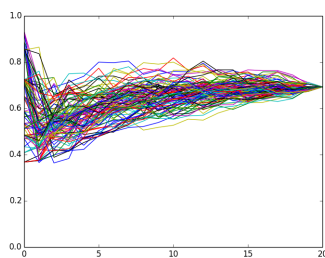
4.2 Local Error Rates

4.3 Block Error Rates

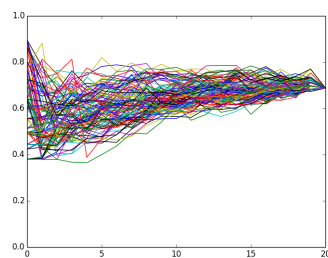
TODO: 1) What is GAINED by blocks in general, 3d blocks in particular? Do the 3d-found groups match the 2d? 2) what am I trying to solve ... 3) what else, honestly? what are my OUTcomes?



(a) A subfigure

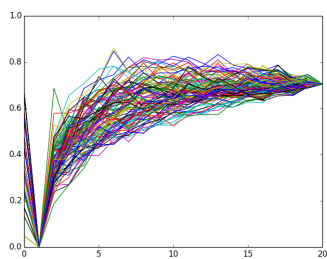


(b) Another subfigure

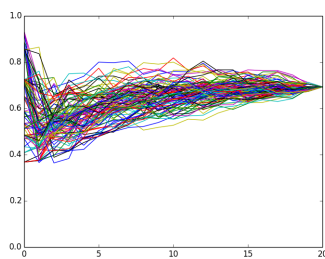


(c) Another subfigure

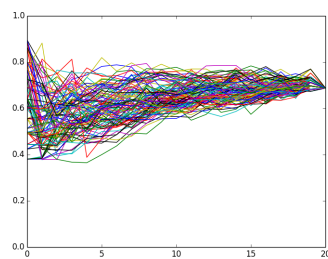
Figure 1: FIRST Row



(a) A subfigure



(b) Another subfigure



(c) Another subfigure

Figure 2: SECOND Row