Siddarth Viswanathan

Study of Variance Reduction in Bipartite Experiments

1. Summary

Bipartite experiments arise often in social theory, economics, and legal analysis. In bipartite experiments there are connections between two types of nodes, which are intuitively referred to as institution nodes (I-nodes) and person nodes (P-nodes). An outcome of interest is defined over the P-nodes and a cause is imposed randomly and measured over the I-nodes. An important contribution to variance reduction is in clustering the I-nodes and imposing the cause randomly at the level of clusters instead of the level of nodes. Correctly clustering the I-nodes can greatly reduce variance if the clustering is performed to maximize the variation among the P-nodes connected to each I-node. Intuitively since each I-node is connected with weights to many P-nodes we can compare these vectors of P-node weights for each I-node and construct a measure which is maximized when the variation among the P-node weights is maximized. Using a simple heuristic algorithm for variance reduction by correlation clustering we implement the algorithm and study the convergence properties of the number of clusters and the objective function.

2. Variance Reduction Algorithm

The purpose of correlation clustering over I-nodes is to reduce the variance of the measurement of the imposed cause by maximizing the variation of the coverage of those I-nodes imposed with the cause. Pouget-Abadie et al. (2018) show using a

standard argument in graph theory that the variance-maximization objective of a bipartite experiment can be rewritten as

$$\Delta = \alpha + \beta \left(\sum_{C} \sum_{j,k \in C} W^{+}_{jk} - \sum_{C \neq C'} \sum_{j \in C', k \in C} W^{-}_{jk} \right)$$

where the *W* matrix is formed from $W_{jk} = \langle \vec{w}_{\cdot j}, \vec{w}_{\cdot k} \rangle - {}^{1}/_{N} \langle \vec{w}_{\cdot j}, \vec{1} \rangle \langle \vec{w}_{\cdot k}, \vec{1} \rangle$ where each vector *w* contains the fundamental edge weights connecting the I-nodes and P-nodes. Intuitively this *W* matrix measures how similar the P-nodes are among every pair of I-nodes. The variance-maximization objective contains constants $\alpha = p^2 - \beta \sum_{j,k} W^{-}_{jk}$ and $\beta = {}^{p(1-p)}/_{N}$ which are constants with respect to each clustering. The variance-maximization objective will be maximized when the *W* values are most positive within a clustering and most negative outside a clustering. Using this criteria we can construct many different clusters to see when the objective is maximized. The interpretation of the final cluster is the cluster design over which to impose the cause to measure with least variance the causal estimand.

Like other constrained graph clustering problems the correlation clustering maximization method is NP-hard and difficult to approximate (Bansal et al., 2002; Charikar et al., 2003; Ailon et al., 2008; Demaine et al., 2006). But Pouget-Abadie et al. (2018) show that a simple local search procedure converges to maximizing the objective. The procedure is to begin with singleton clusters and to move a node from one cluster to another if the objective improves or to merge two existing clusters if the objective improves.

3. Steps to Algorithm Implementation

There are four steps to implement the algorithm for maximizing the objective above. First, generate a *W* matrix summarizing the relationship between I-nodes and P-nodes

of the bipartite experiments. Second, beginning with singleton clusters at each step find out all the possible new clusters which can be gathered by moving a node from one cluster to another or merging two existing clusters. Third, calculate the objective over each possible cluster and select the cluster which maximizes the objective to be the starting cluster for the next step. Fourth, repeat until convergence and observe the final number of clusters.

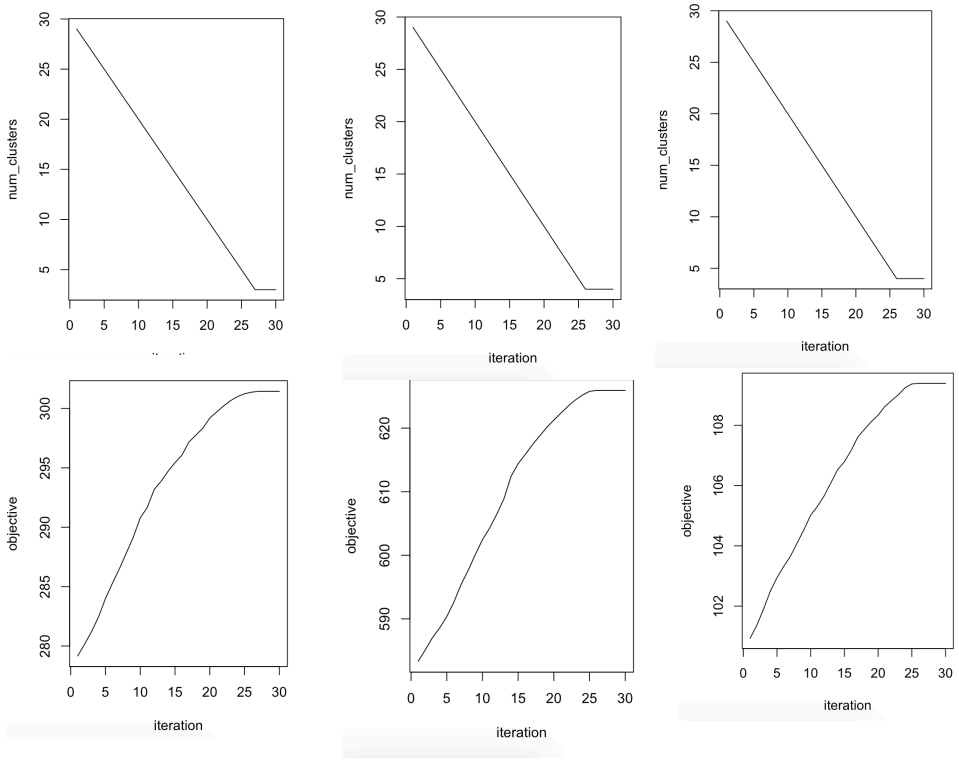
To generate the *W* matrix we first need to define over which P-nodes each I-node is defined over. A simple model here is a Bernoulli outcome for each P-node with the parameter representing the sparsity of influence of each I-node. Now given which P-nodes each I-node is defined over in the bipartite experiment we can determine the distribution governing the weights of the connections of each I-node to the P-nodes. Any continuous or discrete distribution can govern these weights.

The key observation to clarify the space of possible new clusters of I-nodes with which to improve the variance reduction is that if the clusters are represented by distinct natural numbers then a possible new cluster is a cluster with the same cardinality as the original cluster and identical values giving clustered nodes. So the new clusters possible by moving a vertex to a new cluster are all the unique clusters arising from switching the unit value to another existing unit value, and all the new clusters possible by moving an entire cluster to a new cluster are all the unique clusters arising from changing all unit values within a cluster to another existing unit value. After calculating the objective over this space of possible clusters we choose the cluster which maximizes the objective and repeat until convergence.

4. Study of Algorithm

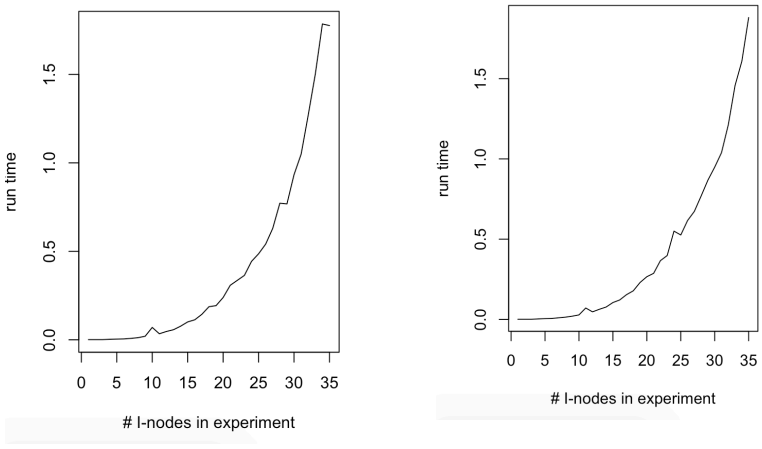
To understand further the algorithm for correlation clustering for variance reduction in bipartite experiments we present studies for 1) how the final number of clusters and convergence properties of the objective respond to changes in the weight distribution 2) algorithm completion time for a fixed number of iterations as the number of I-nodes and P-nodes increases 3) number of iterations to algorithm convergence as the number of I-nodes and P-nodes increases 4) how sparsity aspects of the weight distribution between I-nodes and P-nodes relates to the final number of clusters and objective convergence.

Suppose the weight distribution between 30 I-nodes and 500 P-nodes is governed by a gamma(1,1) distribution, then a uniform(5,10), then normal(0,1). The six figures here show the objective convergence and number of clusters over the algorithm iterations for each of the three generating distributions from left to right. We note that the results are equivalent independent of the number of P-nodes, only the I-nodes governs convergence and number of final clusters properties since the clustering is performed only over the I-nodes.



These six figures reveal that regardless of the generating weight distribution the final number of clusters is around five after the 25th iteration and the objective converges around the 25th iteration. Further clarity is needed as to why convergence seems unaffected by the governing distribution.

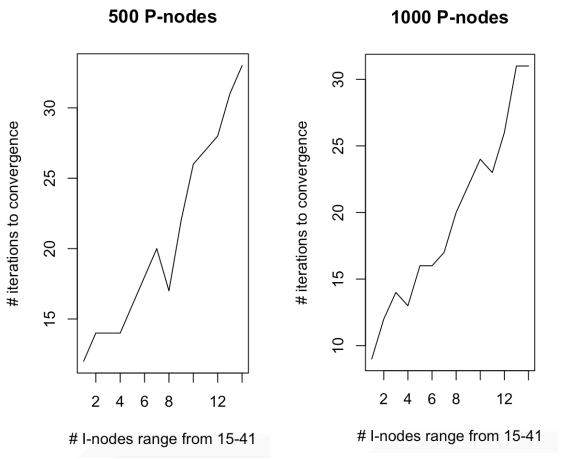
When using the algorithm the most important factor governing runtime is the number of I-nodes. Because the *W* matrix is calculated using weights from the I-nodes and the clustering is done over the I-nodes the number of I-nodes is the main factor governing the algorithm runtime. The following two plots show for a sequence of I-nodes ranging from 1 I-node to 35 and for 500 P-nodes on right and 1000 P-nodes on left the runtime as the number of I-nodes increases.



There is no noticeable effect of the number of P-nodes on the runtime. Importantly we observe an exponential growth in the runtime until a fixed number of iterations as the number of I-nodes increases. Since even with around 30 I-nodes there are iteration runtimes around one second long it is difficult to perform more realistic simulations with thousands of I-nodes without improved computation power. The reason for this exponential increase in runtime is that as the number of I-nodes increases the size of

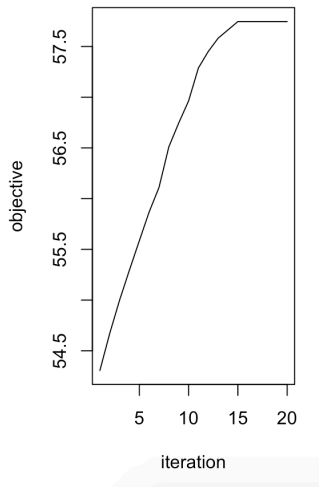
the space increases exponentially of possible new clusters among which we must choose a variance maximizing clustering.

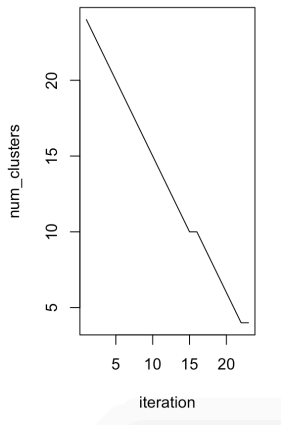
Now letting the I-nodes range in number from 5 to 41 and settings 500 P-nodes on left and 1000 P-nodes on right the following plots show the number of algorithm iterations until convergence of the objective.



Again the number of P-nodes does not have any impact on the number of iterations until convergence. Though it seems that there is a strong linear relationship between the number of iterations until convergence and the number of I-nodes it is possible that for much larger numbers of I-nodes the relationship becomes exponential. Again this growth of number of iterations is a consequence of the exponentially larger possible new cluster size as the number of I-nodes increases which slows the algorithm since each iteration must compare among more possible potential clusterings.

A final question is whether the number of clusterings produced by the algorithm changes significantly if sparsity considerations are introduced in the weight generation between I-nodes and P-nodes. Allowing each I-node a maximum sparsity of 20% over all P-nodes we use a Bernoulli distribution with a uniform(0, .2) probability to govern which P-nodes the I-node is connected to. This models variation among I-nodes with sparse connections to P-nodes. The following figures show the convergence of the objective and number of clusters for 30 I-nodes and 500 P-nodes using the 20% sparsity method.





The number of I-nodes and P-nodes is the same for these figures as in the first figures of this section therefore we would attribute any difference in the iteration number at convergence of objective or number of clusters only to the additional modeling of sparsity. The number of clusters with sparsity converges at the same iteration as without sparsity, while the objective with sparsity converges much before the iteration where the objective converges without sparsity. This quicker convergence to the objective makes sense since each iteration step will be able to adjust for more variance among the I-nodes since the overlaps among the I-nodes are more sparse.

5. Conclusion

Bipartite experiments are a useful formalization for many problems in the social, economic, and legal sciences. These types of experiments capture the natural relation between the societal institutions and the persons these institutions effect in a way which motivates the institutional imagination to measure new imposed causes which

change the societal institutions. A crucial aspect of useful bipartite experimentation is correlation clustering to reduce causal estimator variance. This estimator variance is reduced when variation among the P-nodes influenced by the I-nodes is maximized. We have implemented an algorithm to maximize this empirical variance of P-nodes and have studied convergence and runtime aspects of the algorithm in relation to governing distribution, number of I-nodes and P-nodes, and sparsity.

6. References

[1] Nir Ailon, Moses Charikar, and Alantha Newman. 2008. Aggregating inconsistent information: Ranking and clustering. *J. ACM* 55, 5 (2008), 23:1–23:27.

[2] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. 2002. Correlation Clustering. In *43rd Sympo- sium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*. 238.

[3] Moses Charikar, Venkatesan Guruswami, and Anthony Wirth. 2003. Clustering with Qualitative In- formation. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*. 524–533.

[4] Erik D. Demaine, Dotan Emanuel, Amos Fiat, and Nicole Immorlica. 2006. Correlation clustering in general weighted graphs. *Theor. Comput. Sci.* 361, 2-3 (2006), 172–187.

[5] Pouget-Abadie, J., Aydin, K., Schudy, W., Broderson, K., Mirrokni, V. (2018). Variance reduction in bipartite experiments through correlation clustering. *32nd conference on neural information processing systems*.