1 Assignment of departments

To empirically determine which departments exhibit traits similar enough to be grouped together, we use a maximum likelihood estimation framework. We start by compiling every observed placement of applicants under a specific position, in our case assistant professors of economics and related fields. Each placement is composed of a movement from the department the applicant graduated from toward the organization or department the applicant was hired at. As in traditional group-based likelihood models (Peixoto, 2012) (more citations here?), the set of all such placements forms a directed multigraph, where in our case the vertices are each department in the market and each edge is a different placement between departments. Let A be the adjacency matrix which represents this graph such that $A_{i,j}$ is the number of applicants who graduated from department i and were subsequently hired at department j. Denote N as the total number of departments in the market.

To endogenously account for cases where applicants are hired at positions other than the one being analyzed, we introduce the concept of a *sink*: an organization or department that does not produce any further graduates. While we primarily analyze the market for assistant professors that are hired at Ph.D.-granting departments, each individual applicant has the potential to be hired anywhere; in particular, being hired as an assistant professor is one of many possible employment streams upon graduation with a Ph.D., the other streams including but not limited to: government, private sector and sessional lecturer appointments. As each alternative stream is not involved in the education and support of future Ph.D. graduates, we consider these to be terminal departments. Mathematically, if i is a sink department, then $\forall j \in [1, N], A_{i,j} = 0$.

We define four categories of sinks for the purposes of our estimation, based on those which were observed in practice. The first of these is departments containing alternative academic placements, denoted as Other Academic, which is composed of sessional lecturers and other placements which would not be found on the usual tenure-track stream. In particular, we do not include post-doctoral researchers as this position is often a precursor to other positions in each possible employment stream, and we often observed applicants hired directly to these streams without having completed a post-doctoral position. Additionally, each department included in this sink is considered to be independent from those in the assistant professor grouping, even if a department exists in both, which accounts for the difference in market structure between the hiring streams.

The second sink category, denoted as Government, includes any organization involved primarily in government functions rather than academic functions. These include central banks, government advisors, among other positions.

The third category, denoted as Private Sector, includes any organization operating in the private sector. While many of the placements observed were to economic positions in private sector organizations, this category also includes organizations to which graduates were hired for other purposes. In some cases, the organization was a start-up created by said graduate.

The fourth and final category, denoted as Teaching Universities, is composed exclusively of academic departments that do not graduate Ph.D.s but do hire assistant professors. In particular, we count these departments separate from both the Other Academic sink and the primary groupings, even if they exist in more than one place. This category acts as a sink for assistant professors, a case which is not captured by either of the other three sink categories.

Because a sink department is physically incapable of producing Ph.D. graduates, we consider placements in A from these departments to be impossible rather than having a placement rate of zero. Thus, let M be the number of non-sink departments and denote $\hat{A} = A[1, N; 1, M]$ to be the adjacency matrix which disregards these impossible placement paths.

As in Karrer and Newman (2010), we treat each the number of placements in $\hat{A}$ between two particular departments as a random draw from a Poisson random variable. Since the number of department to department pairs far exceeds the number of applicants in the job market over time, we can safely estimate any particular manifestation of job placements between departments as being a rare, independent occurrence, satisfying the requirements of using the Poisson distribution.

Define t_i to be the index of the true group to which department i belongs to, which we henceforth denote as the *type* of department i — at this stage, we do not decide on the number of such true types, but denote this as K. Then $\hat{A}_{i,j} \sim Pois(\lambda_{t_i,t_j})$. In other words, each job placement is a random draw from a specific Poisson mean associated with the pair of types t_i and t_j. Given K types, there are a total of $K(K+4)$ different possible pairs and thus Poisson means, since there are four additional types associated with our four sinks.

Similar to traditional models (re-cite the same ones?), we then denote the *likelihood* of a particular instance of $\hat{A}$ having been generated by the $K(K+4)$ Poisson means as:

$$\mathcal{L}(\lambda_{t_i,t_j,\forall i,j}) = \prod_{i=1}^{M} \prod_{j=1}^{N} \Pr(Pois(\lambda_{t_i,t_j}) = \hat{A}_{i,j})) \tag{1}$$

In other words, the likelihood of any one job market network being observed is the joint probability that each set of placements between two specific departments occurred. To make computations more readable, as in traditional models (cite the same ones here too?), we compute the log-likelihood as:

$$log(\mathcal{L}(\lambda_{t_i,t_j},\forall i,j)) = \sum_{i=1}^{M}\sum_{j=1}^{N} log(\Pr(Pois(\lambda_{t_i,t_j}) = \hat{A}_{i,j})) \tag{2}$$

The goal, then, is to solve $\max_{\lambda_{t_i,t_j},\forall i,j} log(\mathcal{L}(\lambda_{t_i,t_j},\forall i,j))$, which produces the set of Poisson means that are most likely to have generated the observed job market network. To do this, we use a maximum likelihood approach similar to that outlined in Peixoto (2014).

We start by taking all departments $i \in [1, M]$ and assigning them uniformly at random a $t_i \in [1, K]$. For departments in the four sink categories, we assign them types $K+1, K+2, K+3$ and $K+4$, respectively. The sink types remain fixed throughout the estimation.

Using this random assignment, we then compute λ_{t_i,t_j} as the mean number of placements between an institution with type t_i and an institution with type t_j. That is, for $t_1 \in [1, K], t_2 \in [1, K+4]$:

$$\lambda_{t_1,t_2} = \frac{\sum_{i=1,t_i=t_1}^{M} \sum_{j=1,t_j=t_2}^{N} \hat{A}_{i,j}}{|\{i \mid t_i = t_1\}| * |\{j \mid t_j = t_2\}|} \tag{3}$$

Given these Poisson means for the current allocation, we compute the likelihood of the network having been generated by these means as $log(\mathcal{L}(\lambda_{t_i,t_j},\forall i,j))$. This forms the base likelihood of the random allocation.

We then conduct a modified version of what is known as a Markov Chain Monte Carlo (MCMC) adjustment sequence in traditional models (Peixoto, 2014). Here, we randomly choose between one and three departments to reassign types to, such that the likelihood of the network given recomputed means may be increased. The specific number of institutions to reassign is empirically-determined, and it was observed that varying the number controls the speed at which the estimator converges to completion.

Formally, let $D \in [1, 3]$ be the number of departments being reassigned, and collect $\{i_k, k \in [1, D]\}$ as a set of D random departments. Then for each i_k that currently has type t_{i_k}, we assign a new type $t'_{i_k} \in [1, K] \neq t_{i_k}$.

After assigning new types to the D departments, we recompute the log-likelihood. If the new likelihood value is greater than the current baseline, we let the new value be the new baseline and adapt the type reassignment into the current allocation of types. If the new value is not greater, we ignore the reassignment, maintain the existing baseline, and repeat the MCMC adjustment with a new random choice of departments.

By taking each improvement as given, we eventually converge to a point where no further improvements are possible as the probability of the network being observed cannot exceed 1. Once this point is reached, the maximum likelihood estimate is obtained, as are the Poisson means which support the estimate. The assignment of types to the departments at this maximal level then represents the most likely grouping of departments together based on having placement rates as close to the group means as possible.

In practice, the estimator does not converge to 1, but rather converges to a smaller region. As the network estimate is based on a sample of data rather than a population, it is impossible for the estimator to determine the exact type assignment, but as an estimate, the goal is simply an approximation. In particular, Bickel et al. (2013) demonstrate that maximum likelihood exhibits asymptotic normality, and is thus a consistent estimator of the true parameters which define the job market.

Finally, the maximum likelihood procedure as above requires the number of types K to be fixed as otherwise we obtain a different type assignment for each possible value of K. To pick a particular number of types, we evaluate the estimator at various values of K factoring a penalty term defined in Peixoto (2013) as description length:

$$\mathcal{D} = \frac{K(K+1)}{2} ln(\sum_{i,j} \hat{A}_{i,j}) + N ln(K) \tag{4}$$

such that the new likelihood metric becomes:

$$\mathcal{L}' = log(\mathcal{L}(\lambda_{t_i,t_j,\forall i,j}))) - \mathcal{D} \tag{5}$$

Likelihood is an increasing function of K, as given more parameters to describe the network, the resulting estimate would more closely reflect the true market. In particular, if each department to department pair had its own Poisson mean, the probability of the network being observed would be 1. Description length is also an increasing function of K, since increasing the number of types requires the true network to have an inherent higher complexity, which is less likely to occur. Thus, by taking the new likelihood as the base likelihood minus the description length, there exists a value of K such that $\mathcal{L}'$ is maximized with network complexity taken into account.

2 References

B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1), 2011. doi: 10.1103/PhysRevE.83.016107. URL: https://arxiv.org/abs/1008.3926

P. Bickel et al. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, 41(4):1922–1943, 2013. doi: 10.1214/13-AOS1124. URL: https://arxiv.org/abs/1207.0865

T. P. Peixoto. Entropy of stochastic blockmodel ensembles. *Physical Review E*, 85(5), 2012. doi: 10.1103/PhysRevE.85.056122. URL: https://arxiv.org/abs/1112.6028

T. P. Peixoto. Parsimonious module inference in large networks. *Physical Review Letters*, 110(14), 2013. doi: 10.1103/PhysRevLett.110.148701. URL: https://arxiv.org/abs/1212.4794

T. P. Peixoto. Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models. *Physical Review E*, 89(1), 2014. doi: 10.1103/PhysRevE.89.012804. URL: https://arxiv.org/abs/1310.4378

should also acknowledge Sam Gyetvay's notes, etc