1.2.2 GFP and HST Imagery

Goddard Fabry-Perot (GFP) images obtained by this author and Carol Grady at APO on the 15th of May 2007 are presented. These data were already published in A. Kospal et. al. 2008 [2]. Images were obtained both in H-Alpha at 6563 Angstroms and at 6590 Angstroms. By subtracting the o-band images from the H-Alpha image we were able to image the HH-knots near the star. Imagery from the Hubble Space Telescope (HST), previously published in the same paper is also included as a useful gure. I refer you to A. Kospal et. al. 2008 [2] for more details of the data reduction on both. (The pre-print is available at http://arxiv.org/abs/0710.1431)

1.3 Results

1.3.1 HST

2 Useful Texts

1. Solid State Physics, Ashcroft and Mermin
2. Principles of Magnetic Resonance, Charles P. Slichter
3. Semiconductors and Semimetals Volume 21, J. David Cohen and others
4. Semiconductor Devices Physics and Technology, S.M Sze

References

[1] Quanz S., Henning T., Bouwman J., van Boekel R., Juhasz A., Linz H.,Pontoppidan K. M., Lahuis F., 2007, ApJ, 668, 359

[2] A. Kospal, P. Abraham, D. Apai, D.R. Ardila, C. Grady, Th. Henning, A. Juhasz, D.W. Miller and A. Moor, 2008 MNRAS, 383, 1015-1028

Consumer Basket Analysis and Expected Co-Occurance of a Bipartite Graph

Natalie Bowlus

Swarthmore College
500 College Ave.
Swarthmore, PA 19081
www.swarthmore.edu
natalie.bowlus@gmail.com

Eötvös Loránd University
Pázmány Péter sétány 1/C
Budapest, Hungary
www.elte.hu
Adviser: Kati Vesztergombi

This year, while a student in ELTE's MSc in Mathematics program, I was involved in a project with a post-doctoral fellow in which we attempted to find a theoretical approach to the problem of the expected co-occurrence of two vertices of a bipartite graph. This paper details our work so far and our future projects, as well as my experiences as a master's student and how they are related to the university's larger transition under the Bologna Process.

1. Introduction

For this past year, I have been enrolled in a Master's program in Mathematics at Eötvös Loránd University (ELTE) in Budapest. While this in and of itself might not be extraordinary, my situation was unique in that I was enrolled in a program which did not exist until this fall. Rather, the English-language master's program, of which I was a part, has been created as part of ELTE's transition away from the traditional Hungarian five-year undergraduate degree and towards a more "European" system.

Initially, I had thought to focus my Fulbright research on this transition, and specifically how it impacted foreign-language programs. Through conversations with students and faculty in a variety of English-language programs, as well as my own experience, I have gained an understanding of the concerns of faculty members involved in programs, students who are enrolled, and the challenges faced in terms of attracting foreign students and providing them with an education. Most programs are also in a particular limbo period: that is, ELTE will not have actual Hungarian master's students until next year.

However, many of the questions I had posed myself before arriving turned out to be less interesting than I might have thought. To a large extent, the University's transition is bureaucracy, and most of the programs were not nearly as developed as I had initially believed. While taking a course with a German post-doc in the Computational Biology department, I joined a research project concerning the expected co-occurrence of vertices of a bipartite graph, and specifically how to characterize it in a theoretical manner. This research has become much more professionally and intellectually fulfilling than my original plans. Through this work I learned important programming skills, gained an opportunity to publish and perhaps even opened the door to a job in the near future.

2. Expected Co-Occurrence and the Netflix Prize

2.1. Background

I initially came to this project through a happy coincidence that I attribute to both my visibility as a foreign student at ELTE and the difficulties associated with designing my curriculum. Accidental though it might have been, this research has become one of the most interesting and formational aspects of my year in Hungary. Not only has it opened a door to computer science, a field I was interested in before my arrival but in which I have had little experience; through this research, I have gained the opportunity to publish an academic paper, as well as learned valuable skills which should open up career prospects in the near future.

Nina Zweig, a German post-doctorate student in the Computational Biology Department, was offering a class in the fall broadly targeted at mathematicians, computer scientists and physicists; furthermore, it was being offered in English. In short, it seemed a match made in heaven. One thing led to another, and I was offered to take part in a project resting at the intersection of discrete mathematics and computer science.

Our problem of determining the expected co-occurrence of two vertices stems from the Netflix prize, and is best explained in this context.

Netflix, the well-known online move rental company, maintains a database of users and the films they have rented. The prize itself – a cool million dollars – is offered to the team or individual who manages to develop a more accurate means of predicting how much someone will enjoy a movie based on their prior rental choices. Currently, this is done by looking at other users who have rented the same movies as the user in question.

However, the question of expected co-occurrence appears in many other fields as well, including economics, biology, and pure mathematics – generally speaking, whenever there are two sets of objects (renters and films, bird species and habitats) and scientists curious over whether two objects co-occurring (birds sharing a habitat, for example) is statistically significant. The current formula for determining the expected co-occurrence of two objects, known as the independence model, is frequently far-off from data achieved in an experimental setting, although this does not prevent it from being widely used. Its flaw is that it assumes that the co-occurrence of any two vertices is based only on their respective degrees.

It is easy to construct a counter-example, albeit not one that is likely to be found in real-life. Our goal was to show first that the independence model is almost always incorrect, and then to characterize the difference between the independence model's prediction and the actual expected co-occurrence of any two vertices in a graph. We hoped to find both a theoretical solution and compelling experimental evidence. Although a complete and definitely theory currently remains out of our grasp, we have hope that experimental data will ultimately support our hypothesis.

2.2. Definitions

The most efficient way of handling the data we are interested in is by considering it as a bipartite graph. A *graph* is a collection of vertices and edges. It can be represented visually in such a way that vertices are drawn as points, while edges are lines which connect two vertices. This representation will be used later in the text. A second useful way of representing a graph is known as the *adjacency matrix representation*. Here the graph is depicted as a matrix in which the rows and columns are labeled by the vertices of the graph. The $\boldsymbol{a}_{i,j}$ entry of the matrix is 1 if vertex v_i is connected by an edge to vertex v_j and 0 if not. Adjacency matrices have the advantage of being easily stored and manipulated by a computer.

For a given vertex of the graph v, the *neighbors* of v are the vertices which are connected to it by an edge. If vw is an edge in a graph $\mathbf{G}$, we say that v and w are *adjacent*. In the adjacency matrix representation of a graph, the neighbors of v_i correspond to the set of columns which contain a 1 in row i. The *degree* of a given vertex is the number of neighbors it has. We can also talk about the *degree sequence* of a graph, which is a list in non-decreasing order of the degrees of all the vertices in the graph. For example, {3,2,2,1} is a possible degree sequence for at least one graph. It is important to note that degree sequences are generally not unique; that is, many graphs can have identical degree sequences.

For this project, we consider a special class of graphs, known as *bipartite graphs.* What makes these graphs special is that their vertex set can be partitioned into two classes in such a way that all edges connect vertices from different classes. For example, the Netflix data can be stored as a bipartite graph, where one class of vertices represents the users and the other, the films they check out. An edge connects each user to every film he or she has rented. When thought of this way, it would be silly for an edge to connect two users or two films.

From here on out, we will use $L=\{l_1,\ldots,l_n\}$ and $R=\{r_1,\ldots,r_n\}$ to denote the vertex sets of a bipartite graph. Except where it might lead to confusion, L and R will also denote degree sequences of a bipartite graph. The number of vertices in L and R will be denoted by l and r, respectively.

Given a degree sequence, a graph is *feasible* if it has the given degree sequences and has no self-loops (where a vertex is connected to itself) and multiple edges (i.e., when two vertices v and w are connected to each other by two or more edges). Let $\mathbb{G}(L,R)$ be the set of all feasible graphs with fixed degree sequences L and R, that is, the set of all graphs with the same given degree sequence. Let $\mathbf{G}$ denote an instance of a graph in $\mathbb{G}$ and $E_{\mathbf{G}}$ denote the edge set of a given graph $\mathbf{G}$. $\mathbb{G}(L,R)$ is feasible if there exists at least one bipartite graph $\mathbf{G}$ without multiple edges and with degree sequences L and R.

Let $N(v)$ denote the set of neighbors of vertex v. For any two vertices v_i and v_j, their *co-occurrence*, denoted $co(v_i, v_j)$ is defined as: $co(v_i, v_j)= |N(v_i) \cap N(v_j)|$, i.e., the number of their common neighbors.

Thus, if any graph $\mathbf{G}$ in $\mathbb{G}(L,R)$ is drawn uniformly at random, the *expected co-occurrence* $Eco(v_i, v_j)$ of v_i and v_j is given by:

$$\frac{\sum_{G\in\mathbb{G}} \text{number of co-occurrances of } v_i \text{ and } v_j}{|G(L,R)|}$$

The heart of our theoretical approach concerns the function we have called *move*, which is defined as follows. Given vertices v and w in L and x and y in R such that the degree of x is greater than or equal to the degree of y, which is greater than or equal to 1, if vx and wy are in $E_{\mathbf{G}}$ and wx is not, then delete the edge wy and replace it with wx. We call x and y a *moveable* vertex pair. The general idea behind swap is to maintain the number of edges in the graph while moving edges from vertices with low degrees to those with higher degrees.

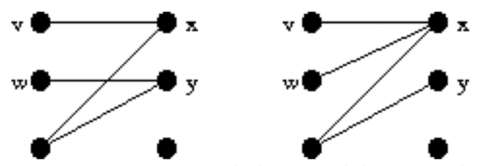


FIGURE 1. MOVE AND THE RESULTING GRAPH.

Under *move* the degree of x is increased by one and the degree of y is decreased by one. The degrees of other vertices in L and R are unchanged. For any feasible move, a graph $\mathbf{G}$ in $\mathbb{G}(L,R)$ is mapped to a graph $\mathbf{G'}$ in $\mathbb{G}(L,R')$, where R' is the degree sequence corresponding to R after moving an edge from y to x.

2.3. The Faultiness of the Independence Model

The key assumption behind the independence model is that a vertex L is equally likely to be connected to any vertex in R; that is,

$$P(l_i, r_s \in E_G) = \frac{\deg(l_i)}{r}.$$

Thus the probability of any two vertices in L sharing a common neighbor in R is given by the sum

$$\sum_{r\in R} \frac{\deg(v)\deg(w)}{r^2} = \frac{\deg(v)\deg(w)}{r}.$$

Therefore, to sum over all possible pairs *(v,w)* L requires the double sum

$$\sum_{r\in R}\sum_{(v,w)\in L\times L} \frac{\deg(v)\deg(w)}{r}.$$

Note that in our case the pair *(v,w)* is identical to the pair *(w,v)*, and we only want to count each pair once.

As the above equations illustrate, the assumptions made by the independence model makes calculating the probability that any two vertices co-occur rather straightforward; hence its popularity in the literature. Generally speaking, calculating the "actual" expected co-occurrence is substantially more difficult. However, in the special case in which all vertices in L have the same degree and all vertices in R have the same degree (note that it needn't be true that $\deg(l)=\deg(r)$), we can explicitly calculate the expected co-occurrence of all pairs (v_i,v_j). In this case,

$$Eco(v_i, v_j) = \sum_{v_i\in L}\sum_{\substack{v_j\in L\\ v_j\neq v_i}} co(v_i, v_j) = \sum_{r_s\in R}\binom{r_s}{2} = r\cdot\binom{r_s}{2}.$$

For the independence model to be an accurate predictor of expected co-occurrence, the sum it predicts should equal the sum we have explicitly calculated; however, this is far from the case. The independence model gives the sum

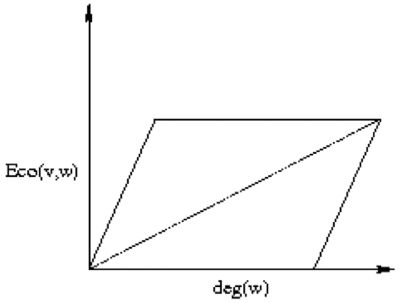
$$\binom{l}{2}\frac{\deg(l)^2}{r} \neq r\cdot\binom{r_s}{2}.$$

From this example we can conclude that the independence model is frequently misleading with respect to the actual expected co-occurrence of any vertex pair (v_i, v_j), although the actually discrepancy between the independence model and the actual expected co-occurrence depends on the specific degree sequences L and R.

2.4. Bounding the Expected Co-Occurrence Function

The expected co-occurrence function has four parameters: $|R|$, $|L|$, $\deg(v)$ and $\deg(w)$ for vertices v and w in L. In order to be able to measure the discrepancy between the independence model and the actual expected co-occurrence for a given graph, in our experiments we will fix $|R|$, $|L|$ and $\deg(v)$ while varying $\deg(w)$. Having done this, we can plot expected co-occurrence on a graph.

FIGURE 2. PICTORIAL REPRESENTATION OF THE BOUNDS ON ECO(V,W), WITH THE INDEPENDENCE MODEL'S PREDICTION ALONG THE DIAGONAL.



Let the x-axis be labeled by deg(w) and the y-axis by the expected co-occurrence of v and w. If we let n=|L|, then the lower bound we have for *Eco(v,w)* is Ω(deg(v),deg(w)) = max{n-deg(v)-deg(w),0}. If *deg(v)+deg(w)* is less than or equal to n, then a feasible graph exists in which v and w have no common neighbors; *deg(v)+deg(w)* since expected co-occurrence cannot be negative, it equals 0. However, if is greater than n, it is necessary that they co-occur at at least |n-*deg(v)-deg(w)*| vertices. The upper bound, on the other hand, is given by O(*deg(v),deg(w)=min{ deg(v),deg(w)}*. It is relatively clear that the number of vertices at which v and w might co-occur is constrained by whichever of the two has smaller degree. The expected co-occurrence predicted by the independence model,

$$\frac{\deg(v)\deg(w)}{|R|}$$

, is on the diagonal of this trapezoid.

Our hypothesis is that for a fixed |R|, |L| and deg(v), if we vary the degree of a vertex w in L, the expected co-occurrence of v and w will remain strictly greater or strictly less than that predicted by the independence model.

2.5. The Move Function: A Preliminary Theoretic Approach

With the *move* function, defined in Section 2.2, we hoped to obtain a theoretical bound on the expected co-occurrence of two vertices, thus demonstrating that it is strictly greater or less than the expected co-occurrence predicted by the independence model. The approach we took was to first prove that *move* is surjective, as shown below. Our next goal was to show that *move* strictly increases the expected co-occurrence of any two vertices. While we could not prove this definitively, we did find arguments which supported this conjecture. Our final step would be to repeatedly use *move* to go from $\mathbb{G}(L,R)$ to $\mathbb{G}(L,R')$, from $\mathbb{G}(L,R')$ to $\mathbb{G}(L,R'')$ and so forth until there are no longer any moveable edges. At this point, we would either explicitly calculate the different between the expected co-occurrence and the independence model's prediction or, more likely, prove that the expected co-occurrence remained strictly greater than or less than the independence model's prediction.

Theorem:

The function *move* from $\mathbb{G}(L,R)$ to $\mathbb{G}(L,R')$ is *surjective*[1].

We would like to show that if there exists a **G** which maps to **G'** in $\mathbb{G}(L,R')$ then for any $\hat{G}$ in $\mathbb{G}(L,R')$, there is some element in ˆ(L,R) which maps to it. Call **G'** the *witness* and let r_i and r_j be the two vertices in **G'** whose degree is different in R' from that in R. Without loss of generality, let the degree of r_i in **G'** be equal to one more than the degree of r_i in **G** and the degree of r_j in **G'** be one less than its degree in **G.** Furthermore, we can say that $r_i > r_j$ in **G'**. All other vertices in R and L have the same degree in both **G**(L,R) and **G**(L,R').

Because the deg of r_i is greater than the degree of r_j in **G**(L,R'), there exists a vertex l such that $l\,r_i$ in $E_{\hat{G}}$ and $l\,r_j$ is not in $E_{\hat{G}}$ for any $\hat{G}$ in $\mathbb{G}(L,R')$. For this reason, we can delete edge $l\,r_i$ and replace it with an edge $l\,r_j$ in order to obtain a graph **G** in $\mathbb{G}(L,R)$. This is indeed a feasible graph because our choice of vertex l insures that we would have no multiple edges. Therefore swap is surjective, as desired. □

The idea behind *move* is that by moving an edge from a lower-degree vertex to a higher-degree one we raise the expected co-occurrence (*Eco*) of many vertex pairs and lower the expected co-occurrence of only a few such the net effect should be that any pair of vertices v and w have a higher *Eco* in $\mathbb{G}(L,R')$ than in $\mathbb{G}(L,R)$.

1 For a function *f: V→W* to be *surjective* means that for every element *w* in *W*, there is some *v* in *V* such that *f(v)=w.*

However, *move* is not without its limitations. First we observe that not every feasible $\mathbb{G}(L,R')$ can be mapped to by the move operation; for example, the graph with degree sequences L and R equal {1,1,1,1}. We also know that if r_i and r_j are vertices which have been acted on by move than then in **G'** the difference between their degrees is greater than or equal to two. Thus any graph **G** such that the difference of deg(r_i) and deg(r_j) is less than or equal to one for all pairs (r_i, r_j) is not mapped to by *move*.

We haven't progressed far enough to know whether or not these drawbacks significantly limit our definition of *move*; it could be that we can provide an accurate enough characterization of the limitations of the independence model without needing to map to such specialized graphs. Another option we have considered is allowing for edges to be swapped away from vertices with degree 1. If we determine that these vertices "disappear" after they lose their edge, then we encounter the problem that the independence model's prediction, which takes into consideration the cardinality of the vertex set R, will change, thus defeating the purpose of our theoretical approach to the problem, described below. It could be that we might continue to count these vertices among the elements of R, although as has already mentioned, we are not in a position yet to say whether or not this would be advantageous.

It is worth noting also that the *move* operation is flexible insofar as different choices of vertex pairs (r_i,r_j) will lead to different feasible families of graphs

$\mathbb{G}(L,R')$. Additionally, two different degree sequences $\mathbb{G}(L,R)^1$ and $\mathbb{G}(L,R)^2$ can map via *move* to the same $\mathbb{G}(L,R')$.

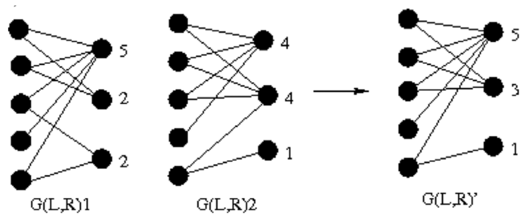


FIGURE 3. TWO DIFFERENT DEGREE SEQUENCES WHICH MAP TO THE SAME SET OF GRAPHS.

Let (r_i,r_j) be a fixed pair of vertices such that *move* decreases the degree of r_j and increases the degree of r_i. Then the cardinality of $\mathbb{G}(L,R')$ shrinks relative to the cardinality of $\mathbb{G}(L,R)$ if there exists a **G** in $\mathbb{G}(L,R)$ such that no moveable edges exist; that is, $N(r_j)$ is a subset of $N(r_i)$. In other words, in the case that every vertex in L which is connected to r_j is also connected to r_i, we cannot move an edge from r_j to r_i without creating a multiple edge and therefore an unfeasible graph. A second circumstance in which the cardinality of $\mathbb{G}(L,R')$ would be smaller than the cardinality of $\mathbb{G}(L,R)$ would be if there were two instances of graphs $\mathbf{G_1}$ and $\mathbf{G_2}$ in $\mathbb{G}(L,R)$ such that they map to the same **G'**.

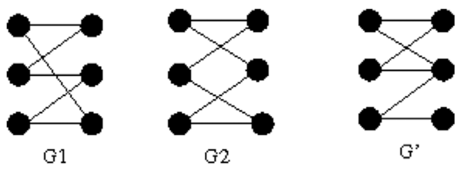


FIGURE 4. TWO DIFFERENT INSTANCES OF GRAPHS WHICH MAP TO THE SAME G'.

With that in mind, let us examine how move affects the *Eco* of any vertex pair. For example, let x be a vertex in L and let r_1 and r_2 be vertices in R such that $x\,r_1$ is an edge in a graph **G** and (r_i,r_j) is a swappable vertex pair. In **G'** $x\,r_1$ is no longer contained in the edge set but $x\,r_2$ is. To understand the expected co-occurrence of any pair of vertices in **G'**, we look at the four possible cases.
Case 1. In the first case, the co-occurrence of any two vertices v_1 and v_2 which are not adjacent to r_1 or r_2 is the same in both **G** and **G'**, since neither is adjacent to a vertex which is increased or decreased in degree.
Case 2. Now examine a vertex v which is connected to both r_1 and r_2. In this case the expected co-occurrence of v with any vertex doesn't change either. For vertices w not equal to x this follows simply from the fact that $N(w)$ is unchanged. The only vertex which might cause us problems is x, but while v and x no longer co-occur at r_1, they do at r_1. Therefore $co(v,x)$ is also the same in **G** and **G'**.
Case 3. For all vertices which are connected to r_2 but not to r_1 their co-occurrence with x is increased by one.
Case 4. The only time the co-occurrence of two vertices decreases is if a vertex v is connected r_1 and not r_2. Here the co-occurrence of v and x is one less in **G'** than in **G**.
However, since there are at most $deg(r_1)\text{-}1$ such vertices, and because of the assumption that the degree of r_1 is less than or equal to the degree of r_2, in a given graph **G** the number of vertices with a decreased degree is strictly less than the number whose degree increases. For this reason, the co-occurrence over all pairs in **G'** is strictly greater than that of **G**. Furthermore, we conjecture that for all vertices v which have a decreased co-occurrence in **G'**, there exists another feasible graph in $\mathbb{G}(L,R')$ where the co-occurrence of v increases correspondingly. Thus in theory, by summing over all feasible graphs in $\mathbb{G}(L,R')$ we would see a net increase in *Eco*.

On the other hand, the independence model's prediction does not change. Our idea is to being with a graph in which explicit computations and the independence model yield the same expected co-occurrence. Next, we would repeatedly use *move* to map $\mathbb{G}(L,R)$ to $\mathbb{G}(L,R')$ to $\mathbb{G}(L,R^n)$, a graph with no move feasible edges. We could calculate the expected co-occurrence of a vertex pair in $\mathbb{G}(L,R^n)$ and in this way obtain an idea of the possible gap between the independence model and the actual expected co-occurrence of vertices in $\mathbb{G}(L,R^n)$.

2.6. Future Avenues of Study

From this point, our theoretical approach reached an impasse. While some work has been done on calculating the number of feasible graphs associated with a given degree sequence, it is enormously difficult. For this reason, although the surjectivity of *move* implies that the cardinality of $\mathbb{G}(L,R')$ is less than or equal to the cardinality of $\mathbb{G}(L,R)$, it is difficult to impossible to explicitly calculate the cardinality of either. Furthermore, although we conjecture that under *move* the expected co-occurrence of a vertex pair (v,w) in $\mathbb{G}(L,R')$ should be greater than or equal to their expected co-occurrence in $\mathbb{G}(L,R)$, we have yet to determine either a rigorous theoretical or computational approach to verifying this.

However, this doesn't settle the issue yet. Instead, in the coming months we hope to change tactics and try for an experimental approach. While at the moment it remains extremely difficult to calculate the expected co-occurrence of any two vertices v and w, it can be estimated. For our experiments, we will use the Netflix data, mentioned in Section 2.1.

First, we will select two vertices and compute their actual co-occurrence; this merely requires examining the adjacency matrix representation of the data. Second, we will generate a large number of random graphs using a Markov chain. A Markov chain is a random walk through the space of all graphs of fixed degree sequences R and L. What the algorithm does is switch edges thousands of times until it obtains a graph that is in theory randomly selected from the set of all feasible graphs. In this newly obtained graph we measure the co-occurrence of the same two vertices v and w. We will do this for approximately 10,000 random graphs, each obtained with a Markov chain starting at our original graph. While this does not amount to a complete sampling of all feasible graphs, it should provide a good approximation. Using this data, we will test our hypothesis that the independence method strictly over- or under-estimates the expected co-occurrence of any two vertices. Whether it is an over- or under-estimation should depend on the degrees of the vertices we select. If the degrees of v and w are significantly greater than average, we predict that the expected co-occurrence should be strictly greater; if their degrees are significantly less than the average, the expected co-occurrence should be strictly smaller; and if their degrees are in the middle, it could go either way.

3. My Experience at ELTE

3.1. Foreign-language Programs and the Bologna Process

Although at first brush the two seem unrelated, my work in the MSc in Mathematics program is intimately tied to the University's general transition under the Bologna Process. I was initially interested in examining this and, more specifically, how it related to foreign-language programs at ELTE, as part of my Fulbright proposal. As it turned out, my research went in a somewhat different direction; however, I think my experiences at the university are indicative of the larger successes and challenges faced by students and faculty at ELTE during this transitional time.

Before 2006, Hungarian higher education was on a 5-3 system, meaning that a students' undergraduate degree lasted five years, after which they received a degree equivalent in most countries to a master's. In contrast, under Bologna these five years are split into a three-year bachelor degree and a two-year master's. In 2006, ELTE ushered in its first class of students who will graduate in the Bologna system. They are set to graduate this summer; the fall will represent the start of the first class of Hungarian master's students.

Last year and now, I saw this transition as a window of opportunity for departments to broaden their programs so as to include English-language master's students. There are a number of reasons it would be desirable for a large research university such as ELTE to attract foreign students. It would increase knowledge of the university abroad, outside of the specific research circles in which it is already known; it would increase the diversity of the student body; it would further cross-cultural education between Hungary and other countries; it would be an opportunity for Hungarian students to practice English, something especially important for those interested in continuing in the natural sciences; and, perhaps most importantly, foreign students' tuition would provide a much-needed source of income.

Before I came to Hungary, I had little idea what to expect regarding the prevalence of ELTE's English-language programs or the number of full-time foreign students at the University. As such, much of my research, especially in the first few months, involved finding out what I didn't know. For example, I knew coming in that the MSc in Mathematics was relatively new, although I didn't realize until I arrived that this was actually only the first year. I also had no idea how large it might be – as it happens, there are currently two students, including myself. The fact of the matter is that the lack of Hungarian master's students puts the state of master's programs in a strange limbo: currently, they have foreign students or none at all. For me at least, this also means that in some sense my program has been in somewhat of a test phase. It is unclear to what extent our curriculum this year will be related to the one used next year for all (Hungarian and foreign) master's students, or whether it will become more flexible.

My final realization, perhaps most important of all, was that the size of master's programs is almost entirely dependent on the individual department. ELTE's website – the main source of information for prospective foreign students – currently offers ten different master's programs in English: Biology, Chemistry, Computer Science, Earth Sciences, English Language and Literature, Mathematics, Physics, Psychology, Social Sciences, and Teaching English as a Second Language. However, some programs have had numerous full-time foreign students for years – most notably the Biology and Psychology programs – while others have merely one or two. And, most surprising to me, some programs are not only without full-time foreign students but also many members of the faculty disinclined to teach in English, perhaps largely due to a lack of financial incentives for them to take on the extra workload. To varying degrees all programs struggle to a greater or lesser extent by the same problems: inadequate advertising, difficulty fitting students into the existing curriculum, and lack of a critical mass of students to make a program financially and practically viable. My experience was also shaped by such concerns.

3.2. My Experiences in Mathematics

Although I had first heard about ELTE through professors who teach there, I found out about the English-language MSc program through browsing the website, much like the other foreign student currently in the program. In fact, almost every student I talked to came to ELTE either randomly or because of personal connections to Hungary, rather than because of knowledge of the university itself. Corresponding with various professors, I was told the math program was very new – only upon arriving did I find out that it is actually in its first year.

Perhaps because the program is advertised largely through the internet, or because we are the first non-Erasmus foreign students, or because we are the first students in the master's program, many of the students I encountered in the first few weeks were unaware that there were foreign (or master's) students at ELTE at all, although they seemed for the most part pleasantly surprised. Some faculty members were also unaware of the existence of a foreign-language master's program.

Like the many of the new master's programs at ELTE, our curriculum is based on the last two years of the old Hungarian degree. To graduate, students are expected to complete three modules in different fields, where a module consists of approximately seven courses. A culminating thesis is also required, although these two alone are not enough to complete the degree – some additional coursework is also required in order to meet the required number of credits. When I was applying for the program, students were given the option of completing the degree in two or three years, although I am not sure if this will still be available to Hungarian students starting next year.

Not every course is offered every year, even in Hungarian. Determining which courses can be offered in English is the job of a professor specifically in charge of foreign students. Although I have heard that a certain number of courses are officially offered in English, practically speaking, if there aren't foreign students in the classes, lectures are in Hungarian. As one professor put it, "It's ridiculous – I speak Hungarian, the students speak Hungarian. Why teach in English?" Whether the course will be taught in English to accommodate a foreign student is a decision left up to the professor and the other students. Should everyone agree, than a class can be taught in English. In my experience in the math department, professors have been generally more than willing to teach in English, for which I am grateful. Through my conversations with other students and professors in various departments, I have come to understand that their coursework consist almost entirely of one-on-one reading courses with professors (in the case of the smaller programs) or courses entirely for English-speaking students (Psychology, for example). In either case, the end result is general isolation from the Hungarian students at the university. I feel that were it not for my lecture courses, meeting Hungarian students at the university would have been substantially more difficult.

However, my curriculum has been supplemented with reading courses, when courses were not offered in a semester when I needed them, or when courses officially in English overlapped. Although I would not chose to have a curriculum composed entirely of reading courses, these one-on-one opportunities with professors have been among the highlights of my studies. I think in very few places would a master's student have the face-to-face time I have with faculty. In my case, reading courses have also served a third purpose: giving me access to material which would, generally speaking, be covered in the Hungarian students' BSc. During this year, I was extremely aware of the fact that the American liberal arts education is radically different from the education my compatriots are receiving in Hungary and other parts of Europe. While at times, explaining my educational background proved an invigorating challenge – I was even invited to give a presentation at the English-language segment of the 2008 Varga Tamás Day Conference, held at ELTE.

On the other hand, I was painfully aware approximately every other week that, while I feel I made the best of my undergraduate education and received a very sound basis in mathematics as well as other subjects, I have less of a background in mathematical than fourth- and fifth-year students my age. This occasionally made it difficult to fill my schedule, and meant even in some cases that I was given a course equivalent to what a Hungarian student might encounter in his or her third year. To some extent, this is a consequence of my personal educational choices, and the differences in the university systems of Hungary and the United States; however, I think it is a problem that will be encountered at ELTE with increasing frequency.

While I have no way of knowing now how many foreign students might attend ELTE's math program in the near future, there will certainly be a huge influx of Hungarian students. Many of them will come from ELTE, meaning faculty can still expect students to know a body of set material; however, many others will come from smaller universities which cannot be expected to maintain full master's programs after the Bologna transition. For these students, it might be necessary to increase the flexibility of the master's programs and, for example, offer courses previously restricted to bachelor's students. It seems for now that although a master's curriculum is currently in place, it will take some years of adjusting to come up with a program which meets students' various needs.

Overall, however, my experience was positive. Although at times navigating a university system primarily in another language was frustrating (in particular, figuring out the convoluted bureaucracy), I received help unstintingly from students and professors. I was actually surprised at times how concerned about my well-being they could be. When I first arrived, I was frequently asked whether or not I had friends, and whether I was settling in alright. Later, I was offered help finding a doctor, calling the BKV to complain about misbehaving controllers, and (most importantly), translating volumes of mathematics or other material into English. I hope for my part that I was also able to contribute to the math department, whether by trying to explain the American healthcare system, talking politics, editing graduate school applications, or just by being a native English speaker.

4. Conclusion

My time in Hungary this year has been overwhelmingly positive. While our research on expected co-occurrence has moved in fits and starts, I am optimistic about our results so far. I believe that Nina and I have several further avenues for study in the near future, as well as a good chance of combining this work with other existing results to produce a conference paper publishable some time in the next year. Furthermore, aside from the aforementioned research, my time as a student at ELTE has been both personally and intellectually fruitful. Although I encountered occasional frustration on account of being one of only a handful of full-time foreign students, these were far outweighed by the benefits: the huge increase in my knowledge, connections I've made with students and professors, as well as the benefits incurred by standing out in a crowd. Regarding the Bologna Process itself, I feel that much of the changes will come down to "only time will tell." It seems unlikely that foreign-language programs will expand dramatically unless more fundamental changes take place; however, hopefully a few years' time will assuage doubts of professors regarding the new curriculum.

I would like to thank several individuals in particular: Kati Vesztergombi, Gyula Károlyi, and Nina Zweig, for advising me in official and unofficial capacities; András Gács and Zoltán Kiraly, for the time they spent attempting to educate me; and the Hungarian-American Fulbright Commission for making this entire opportunity possible.

Bibliography

Eötvös Loránd University. 17 Apr. 2009 <http://www.elte.hu/en>

Gionis A, Mannila H, Mielikäinen T, Tsaparas P (2006) Assessing data mining results via swap randomization. In: Proceeding of the 2006 ACM SIGKDD international conference on knowledge discovery in databases (KDD'06), Philadelphia, PA, pp 167–176

Netflix Prize. Netflix. 17 Apr. 2009 <http://www.netflixprize.com/index>.

Zweig, Nina and Christian Groth. Experimental and theoretical analysis of the expected co-occurrence in market baskets containing single-buy items. Pending Publication.

Romani Education in Hungary: History, Observances and Experiences

Eric Lopez

Hofstra University
Hempstead, NY 11549-1000
www.Hofstra.edu
eric_lopez_82@yahoo.com
Advisor: Dr. Linda Longmire

Central European University
Nádor utca, 09.,
1051Budapest, Hungary
Advisor: Dr. Anton Palinka

As a former advocate for students in Harlem, New York, I arrived in Hungary with a set of expectations regarding the educational injustices faced by the Roma. However, my experiences could not have prepared me for what I would observe and experience during my time there. Researching articles, studies, reports, and statistics has yielded valuable information on the Roma, yet it was my observations at the Dr. Ambedkar School in Sajokaza and Hegymeg that clearly illustrated the challenges and issues faced by students, parents, educators, and advocates in Hungary today. After nearly one year of observation and research on the educational inequalities faced by the Roma in Hungary, one thing has become increasingly clear; Hungary's efforts towards educational equality must rely more on Roma communities and organizations if it is to achieve harmonious integration between its Roma and non-Roma citizens. My observations at the Dr. Ambedkar School in the northern county of Borsod-Abaúj-Zemplén, allowed me the unique opportunity to draw several comparisons between Hungary's efforts towards providing a more inclusive educational environment for its Roma minority and some of the harsh realities faced by its most economically deprived Roma communities.