Candidate Number: 552597

Project Number: TP08

Project Title: Mathematics Genealogies and
The Movements of Academics

Supervisor: Mason A. Porter

Word Count: 6462

Mathematics Genealogies and The Movements of Academics

From the Mathematics Genealogy Project, a database of maths-related PhD theses, I extract the movements of academics between universities on an international scale. I show that the extracted movements are incompatible with several features of a random graph based on the configuration model. Using Bayesian hypothesis testing, I find evidence that an academic is more likely to move after their PhD if their supervisor has moved since their own PhD. I discuss how assortative mixing potentially gives rise to this correlation. Considering only the movements between 187 US universities, I show that academics move, at least indirectly, between every pair of universities. Finally, I explore time dependent trends in the movements and, in two cases, show how they are consistent with historical events.

1 Introduction

There are a number of questions regarding academia for which answers are formed through consensus. An example would be the surveys that underlie many university league tables. Despite the controversy such league tables often entail [1], they have become an indispensable reference for the application process that every student must go through. Bowman and Bastedo have found evidence of league tables significantly affecting where applications are made [2]. Hence, there is motivation to find a more acceptable or, at least, less disputable alternative to consensus.

The Mathematics Genealogy Project (MGP) is a database of completed mathematics PhDs where mathematics is used in the broadest sense to include statistics, computer science, etc. [3]. MGP is discussed in detail in Section 3.1. Researchers have recently used the large volume of data gathered by MGP, as an alternative to consensus, to gain insights into academia. Malmgren, Ottino & Amaral study the role of mentorship in protégé performance through the number of PhD students an academic supervises [4]. Myer, Mucha & Porter apply the method of *hubs and authorities* [5] to the universities found in MGP and establish that authority scores correlate with league table rankings [6].

Academics will occasionally change the institution they are affiliated with. The process of an academic changing institution in their career, from PhD onwards, will be referred to as the academic *moving*. The moves an academic makes will be influenced by numerous considerations, some of which may be unique to the individual. In a sample containing a large number of moves by many academics, the widely important considerations will result in the most significant trends. What considerations are widely important is unclear and it could be that, on a large-scale, movements can be treated as random. If large-scale movements do correlate with certain factors, the underlying causes may still remain uncertain; it is known to be difficult to determine the root of a correlation when there are both viable social and environmental causes [7].

While the subject of this project is not physics in the strictest sense, the topic is one that is likely to be of interest and relevant to a physicist, if not to all with an academic background. The methods of analysis applied to the data are widely transferable to many branches of science. In particular, networks have applications in statistical mechanics [8] and researchers in the field of network theory often have a background in physics [9, 10].

The rest of the report is as follows: In Section 2, I give a background for networks and Bayesian hypothesis testing. In Section 3, I discuss MGP in detail and use the data to extract movements of academics between universities, explaining my procedure. With the movements, I create a network. In Section 4.1, I develop a random graph model and compare it to the movement network obtained from MGP. In Section 4.2, I present evidence, based on Bayesian hypothesis testing techniques [11], that an academic is more likely to have moved at least once in their career to date if their supervisor has moved at least once. I then discuss how such a dependency could be a result of assortativity [12, 5]. In Section 4.3, I investigate the network of movements between 187 US universities; and, in Section 4.4, I examine time-dependent movement trends over the past century, highlighting two universities with trends of particular interest. In Section 5, I conclude and comment on my results.

2 Background

2.1 Network Theory

2.1.1 General Definitions

A *network* is a system containing a number of objects that all have a common defining property and interact in some way. Details are discarded; only the aggregate structure is considered. A *graph* is a representation of a network where interacting objects are depicted as dots (called *nodes*) connected by lines (called *edges*). The definitions found throughout Section 2.1 are based on those found in [5].

Examples of networks include the Internet (nodes are computers and routers, edges represent a connection between them) and social networks (nodes are people, edges represent acquaintance). In physics, networks can arise in statistical mechanics when nearest-neighbour approximations are too weak but infinite range interactions are too strong [8].

In some cases, edges have a direction. For example, in a family tree, the relation of parents to their offspring is not the same as their offspring's relation to them. To indicate the irreversibility of this network, a suitable convention could be to equip each edge with an arrow pointing from the parents to the offspring. These are known as *directed* graphs.

In a graph, a *path* is a sequence of nodes in which there is an edge connecting each node to the next and previous nodes in the sequence. In a directed graph, a path requires that there must be an edge that points from each node in the sequence to the next. For example, in Fig. 2.1, there is a path from Alice to Deb but not from Deb to Alice.

A *cycle* is a path in which the first and last nodes of the path are the same. Family trees are *acyclic*

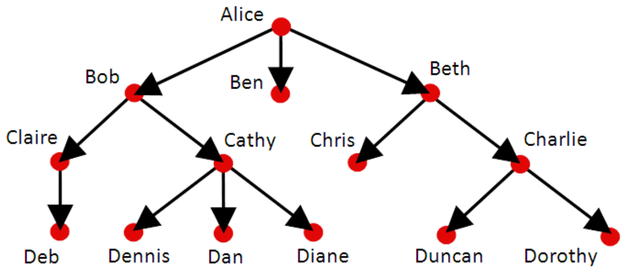

Figure 2.1: A family tree; three generations of Alice's descendants are displayed, spouses are ignored.

in the sense that an edge cannot be directed from a younger generation to an older generation: there are no cycles. Consequently, the graph can be laid out with all arrows pointing down (see Fig. 2.1).

A network can be represented by an *adjacency matrix, A*, where the element A_{ij} gives the number of edges or, more generally, the *weight* of the edge from node i to node j [5]. The total weight of edges from i is the *out-degree*, O_i, and the total weight of edges to i is the *in-degree*, I_i:

$$O_i = \sum_j A_{ij} \ , \ I_i = \sum_j A_{ji} \qquad (2.1)$$

Note that sums over a node label implicitly sum over all nodes.

For an undirected graph, each edge between i and j is considered to be both an edge pointing from i to j and an edge pointing from j to i such that $A_{ij} = A_{ji}$ and the *degree* of i is I_i $(= O_i)$.

2.1.2 Transitivity

The likelihood that any two nodes i and j are connected given both i and j are connected to some node k is called the level of *transitivity* in the network. The *clustering coefficient, C*, is a measure of transitivity [5]. This quantity is most intuitive for undirected graphs with only edges of unit weight or no weight between nodes. In terms of the adjacency matrix, this requires that $A_{ij} \in \{0, 1\}$ and $A_{ij} = A_{ji}$ for all i and j. It is possible to convert a general directed graph to this form, at the cost of losing information, by defining a simplified adjacency matrix,

$$\widetilde{A}_{ij} := \begin{cases} 0 & \text{if } A_{ij} + A_{ji} = 0 \\ 1 & otherwise \end{cases} \qquad (2.2)$$

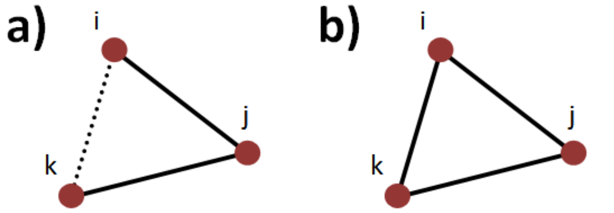

Figure 2.2: a) The connected triple ijk (independent of whether there is an edge between i and k or not). b) The triangle ijk. Note that the triangle contains the three connected triples ijk, jik and ikj.

Once dealing with a graph of this form, we can define what is meant by a *connected triple*. The nodes i, j and k form the connected triple ijk if there is an edge between i and j and an edge between j and k ($\widetilde{A}_{ij} = \widetilde{A}_{jk} = 1$). Note the connected triple ijk could also be denoted kji. If there is a connected triple ijk and an edge between i and k, the nodes i, j and k form a *triangle* ($\widetilde{A}_{ij} = \widetilde{A}_{jk} = \widetilde{A}_{ik} = 1$), see Fig. 2.2. The number of connected triples in a graph is

$$\sum_i \frac{1}{2}\left(\sum_j \widetilde{A}_{ij}\right)\left\{\left(\sum_k \widetilde{A}_{ik}\right) - 1\right\} \tag{2.3}$$

The clustering coefficient is then straightforward to define:

$$C := \frac{(\text{number of triangles}) \times 3}{(\text{number of connected triples})} \tag{2.4}$$

In a similar way, each node can be assigned it's own local clustering coefficient:

$$C_L^i = \frac{\text{number of triangles } ijk}{\text{number of connected triples } jik} \tag{2.5}$$

where the number of triangles ijk is the number of triangles with node i as a vertex and the number of connected triples jik is the number of connected triples with i in the middle given by

$$\frac{1}{2}\left(\sum_j \widetilde{A}_{ij}\right)\left\{\left(\sum_k \widetilde{A}_{ik}\right) - 1\right\} \tag{2.6}$$

Note (2.3) is the sum of (2.6) over all nodes.

The *mean local clustering coefficient*, C_L for the entire network is the arithmetic mean of C_L^i taken over all nodes.

2.1.3 Random Graphs

One can construct graphs that have a common property but are otherwise random. The structural similarities these graphs tend to have will be attributable to their common property. A *random graph model* is defined as a probability distribution over all possible graphs with a particular property [5].

In Section 4.1, a random graph is considered based on the *configuration model* in which the degree of each node is fixed. The configuration model developed here has been adapted from [5] to apply to directed graphs without *self-edges* ($A_{ii} = 0$ for every node i). We begin by specifying the *degree sequence*. That is, for each node

$$I_i = \iota_i \ , \ O_i = \Omega_i \tag{2.7}$$

where ι_i and Ω_i are constants we have freedom to choose. Each node is then given as many *out-stubs* as its out-degree and as many *in-stubs* as its in-degree, see Fig. 2.3.

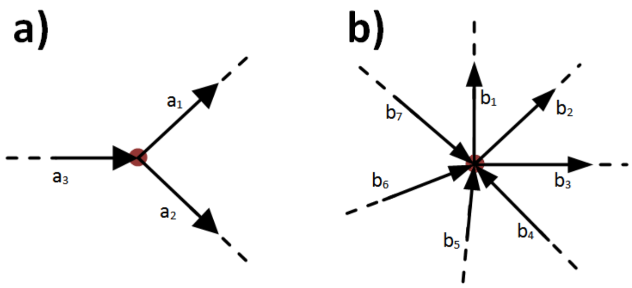

Figure 2.3: a) Node with 2 out-stubs and 1 in-stub b) Node with 4 in-stubs and 3 out-stubs. The out-stubs on one node may be paired in any way with the in-stubs on the other node. For example, a_1 can be connected to b_7.

Every in-stub can be paired with an out-stub on a *different* node. Each unique way of pairing every in-stub with an out-stub is known as a *configuration*. The total number of configurations is ω.

Suppose there is some adjacency matrix, A, that satisfies the constraints $I_i = \iota_i \ , \ O_i = \Omega_i$. Consider an arbitrary configuration corresponding to A, the stubs on every node can be permuted to produce new configurations with the same adjacency matrix, A. However, for each i and j, there will be A_{ij} edges from i to j. Consequently, $A_{ij}!$ of these permutations will give the same configuration. The total number of configurations corresponding to A is [5]

$$\frac{\prod_i \Omega_i! \iota_i!}{\prod_{i,j} (A_{ij}!)} \tag{2.8}$$

If we suppose that each configuration is equally probable, we can divide (2.8) by the total number of configurations, ω, to give the probability of a graph in this configuration model having the adjacency matrix A:

$$pr_{CM}(A) = \frac{1}{\omega} \times \frac{\prod_i \Omega_i! \iota_i!}{\prod_{i,j} (A_{ij}!)} \tag{2.9}$$

Another important result for the configuration model is the expected number of edges from i to j. Denote the total edge weights in the graph by T:

$$T := \sum_{i,j} A_{ij} \tag{2.10}$$

If self-edges were allowed, an out-stub would be equally likely to connect to any in-stub [5]. Hence, for a model with self-edges, the expected number of edges from i to j is given by [5]

$$\langle A_{ij} \rangle = \frac{\Omega_i \iota_j}{T} \tag{2.11}$$

If $\sum_i \langle A_{ii} \rangle \ll T$, the expected number of self-edges is small relative to the total number of edges. Thus, the expected number of edges from i to j in our model without self-edges would approximately be given by (2.11).

2.1.4 Assortative Mixing

In a network, the tendency for nodes to have edges connecting them to other nodes with a common property is called *assortative mixing* [5, 12]. A network in which there is assortative mixing is *assortative.* The *modularity, Q*, for a graph with adjacency matrix A is defined as [13]:

$$Q := \frac{1}{T} \sum_{i,j} \left(A_{ij} - \langle A_{ij} \rangle \right) \delta(prop_i, prop_j) \tag{2.12}$$

where $\delta(prop_i, prop_j)$ is the Kronecker delta and $prop_k$ is some property of node k such that $\delta(prop_i, prop_j) = 1$ if nodes i and j have the same property and 0 otherwise.

The A_{ij} term gives the extent of the tendency for like to connect to like in the graph while the $\langle A_{ij} \rangle$ term gives the extent of the tendency for like to connect to like expected in a random, directed graph with the same degree sequence, see (2.11). The modularity of a graph is thus a measure of the difference between the extent of assortative mixing in the graph and the extent of assortative mixing expected by chance.

2.2 Bayesian Hypothesis Testing

The methods used in this section are based on those in [11]. The following notation is used: $pr(W, X|Y, Z)$ is the probability (or probability density function, evident from context) of W and X given Y and Z.

The elementary operations of probability theory are the *product rule* and *sum rule*:

$$pr(X, Y) = pr(X|Y) pr(Y) \tag{2.13}$$

$$pr(X \text{ or } Y) = pr(X) + pr(Y) \tag{2.14}$$

where X and Y are mutually exclusive in the sum rule (2.14).

From (2.13), *Bayes' theorem* can be derived:

$$pr(X|Y) = \frac{pr(Y|X) pr(X)}{pr(Y)} \tag{2.15}$$

Suppose there is some data, d, from experiments on a system for which we have a hypothesis. To address whether the hypothesis is correct, we must have a *likelihood function*, $pr(d|p)$, for the system. The likelihood function gives the probability of measuring d for the system and depends on a parameter p, which takes some value in the range $p_{min} \leq p \leq p_{max}$.

The hypothesis, Θ – the converse of which is denoted by $\bar{\Theta}$ – claims p can only take a value from a subset of $p_{min} \leq p \leq p_{max}$. A value of p allowed by the hypothesis, will be written as $p \in \Theta$. Using the sum rule (2.14), we can express the probability that Θ is correct by summing the probabilities of $p = \widetilde{p}_i$, given the relevant data, over all $\widetilde{p}_i \in \Theta$:

$$pr(\Theta|d) = \sum_{\widetilde{p}_i} pr(p = \widetilde{p}_i|d) \tag{2.16}$$

Analogously, when the subset contains continuous values of p, we can express the probability that Θ is correct as the integral of the probabilities of $\widetilde{p} \leq p < q + d\widetilde{p}$. In this case, $pr(p = \widetilde{p}|d)$ is now the probability density of $\widetilde{p} \leq p < \widetilde{p} + d\widetilde{p}$ (the equals is retained for notational brevity):

$$pr\left(\Theta|d\right) = \int\limits_{\widetilde{p}\in\Theta} d\widetilde{p}\, pr\left(p = \widetilde{p}|d\right) \tag{2.17}$$

$\widetilde{p} \in \Theta$ is used to imply that the integral is over all values of p that are in agreement with the hypothesis Θ. The above expression is also true when each Θ is replaced by a $\bar{\Theta}$.

As an example, the parameter p could be the probability of obtaining heads by flipping a coin. The likelihood function for the number of heads, h, obtained in H flips would be a binomial distribution. We could compare the probability of the coin being biased to heads ($1/2 \leq p \leq 1$) to the probability of coin being biased to tails ($0 \leq p \leq 1/2$).

We are interested in the ratio of the probability of the hypothesis to the probability of its converse given the data, d:

$$\frac{pr\left(\Theta|d\right)}{pr\left(\bar{\Theta}|d\right)} = \frac{\int_{p_1\in\Theta} dp_1\, pr\left(p = p_1|d\right)}{\int_{p_2\in\bar{\Theta}} dp_2\, pr\left(p = p_2|d\right)} \tag{2.18}$$

To deal with the terms involving $pr\left(p = \widetilde{p}|d\right)$, apply Bayes' theorem (2.15),

$$pr\left(p = \widetilde{p}|d\right) = \frac{pr\left(d|p = \widetilde{p}\right) \times pr\left(p = \widetilde{p}\right)}{pr\left(d\right)} \tag{2.19}$$

If, prior to consideration of the data, we have no basis to believe the hypothesis over its converse, we take

$$pr\left(\Theta\right) = pr\left(\bar{\Theta}\right) \tag{2.20}$$

If, also prior to consideration of the data, we have no reason to believe $p = p_1$ over $p = p_2$ for any $p_1, p_2 \in \Theta$, we take $pr(p = \widetilde{p}) = \rho_\Theta = constant$. With (2.17), this implies

$$pr\left(\Theta\right) = \int\limits_{\widetilde{p}\in\Theta} d\widetilde{p}\, \rho_\Theta \tag{2.21}$$

$$\Longleftrightarrow \rho_\Theta = \frac{pr\left(\Theta\right)}{\int_{\widetilde{p}\in\Theta} d\widetilde{p}} \tag{2.22}$$

with analogous expressions following for $\bar{\Theta}$ when each Θ in the above paragraph is replaced by $\bar{\Theta}$. Application of (2.20) yields

$$\frac{\rho_\Theta}{\rho_{\bar{\Theta}}} = \frac{\int_{s\in\bar{\Theta}} ds}{\int_{t\in\Theta} dt}$$

The ratio in (2.18) is then

$$\frac{pr\left(\Theta|d\right)}{pr\left(\bar{\Theta}|d\right)} = \frac{\int_{p_1\in\Theta} dp_1\, pr\left(d|p{=}p_1\right)}{\int_{p_2\in\bar{\Theta}} dp_2\, pr\left(d|p{=}p_2\right)} \frac{\int_{s\in\bar{\Theta}} ds}{\int_{t\in\Theta} dt} \tag{2.23}$$

This ratio is a measure of how much we believe the hypothesis is correct (given the data, d). A ratio of more than 100 is conventionally taken as conclusive [14].

To return to the example, the probability of getting h heads from H flips will be binomially distributed so we can write

$$pr\left(d|p = \widetilde{p}\right) = \binom{H}{h} \widetilde{p}^h (1 - \widetilde{p})^{(H-h)} \tag{2.24}$$

To answer if the coin is more likely to land heads-up we hence calculate the ratio in (2.23) using (2.24):

$$\frac{pr(p{>}1/2|d)}{pr(p{<}1/2|d)} = \frac{\int_{0.5}^{1} dp_1 (1{-}p_1)^{(H-h)} {p_1}^h}{\int_{0}^{0.5} dp_2 (1{-}p_2)^{(H-h)} {p_2}^h} \tag{2.25}$$

where the integrals on the far right of (2.23) have cancelled.

We could also ask if the coin is fair: $p = 1/2$. Here, instead of applying (2.17), the hypothesis is more simply expressed as $pr(\Theta|d) = pr\left(p = 1/2|d\right)$. We then assign $pr\left(d|p \neq 1/2\right) = pr\left(d|p = 1/2\right)$ as in (2.20).

The ratio of the probability of the coin being unfair to the probability of the coin being fair is

$$\frac{pr\left(p \neq 1/2|d\right)}{pr\left(p = 1/2|d\right)} = 2^H \int\limits_0^1 d\widetilde{p} (1 - \widetilde{p})^{(H-h)} \widetilde{p}^h \tag{2.26}$$

Note that in the limit of the integral, the contribution of $\widetilde{p} = 1/2$ is vanishingly small.

3 Discussion of Data

3.1 Mathematics Genealogy Project

The data used in this project contains information on 137,101 PhDs, which has been obtained from the Mathematics Genealogy Project [3] (MGP). MGP was started by Harry Coonce in 1996 with the aim of compiling a database of one's academic ancestry [15]. The project is run by the North Dakota State University Mathematics Department and hosted online by the American Mathematical Society. The data is gathered from sources such as *Dissertation Abstracts* [15] and through members of the public entering information online. The reliability of information submitted by the public cannot be taken for granted. For example, there are two PhD thesis entered as having been completed in 2012, in the future. The data collected for each PhD are

- The student's name
- The supervisor's name
- A second supervisor's name (if applicable)
- The year of completion
- The university's name

The MGP data constitute an *acyclic directed graph* [5], similar to a family tree, when academics are taken as nodes and an edge from node i to node j indicates that i supervised j. A study on this network has already been published regarding the fecundity of academics [4].

A other networks are also contained in the MGP data. Work is currently being done on the directed network with the universities as nodes [6]. The weight of an edge from university i to university j is taken as the number of academics who, having completed their PhDs at university i, went on to supervise at least one PhD at university j. This network will be referred to as the *supervisor network*.

This project is concerned with another network, which also takes universities as nodes. It is different to the supervisor network in that the weight of an edge from university i to university j represents the number of moves made by academics from i to j. This network will be called the *movement network*. In the case of a single academic, Fig. 3.1 displays the graphs for (a) the movement network and (b) the supervisor network.

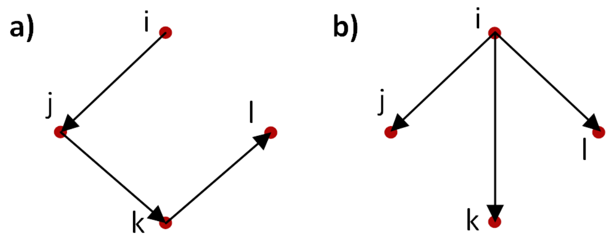

Figure 3.1: a) *Movement network* - the graph for an academic who moves from i to j, j to k and k to l, in that order. b) *Supervisor network* - the graph for an academic who completed their PhD at i before supervising PhDs at j, k and l.

3.2 Extracting Data on Movements

If an academic, labelled a, completed their PhD in year y_1 at university i, then a must have been at i at some point in year y_1. Now suppose a has a student who completes their PhD in year y_2 at university j, then we also know that a was at university j in year y_2. In this way, every student marks where their supervisor was at the date of their PhD completion.

In the example above, a is taken to have moved from i to j in y_2. The move could have occurred prior to y_2. The year given to moves extracted in this way is an upper bound on the year in which the move occurred. The time ordering of an academic's moves is expected to be accurate in all or almost all cases. Clearly, there will be missing moves because an academic does not necessarily have a student at every university with which they have been affiliated.

Of the 137,104 PhD entries contained in the MGP data, 16,070 had no supervisor, 8,291 had no university, and 9,305 had no year. PhD students with a supervisor but without a year or university will necessarily be ignored and so are effectively missing students of their respective supervisors. The first location of an academic who has no PhD university entry is taken as the university at which they first supervise a PhD. Given the culture in the US of academics changing institutions between different stages of their careers [16], the first move made by those from the US without a PhD university entry is likely to be missing. Missing first moves, unlike missing intermediate moves, don't lead to incorrect moves being inferred from the data - an example is illustrated in Fig. 3.2.

Occasionally, an academic will remotely supervise a student at a different university or, after a move, the last student at the old university will complete their PhD after the first student at the new university. In these cases, there is a risk of inferring moves that never occurred. To deal with this, if an academic apparently moved from i to j and returned to i (from j or otherwise) within a five year period, all intermediate moves are ignored, the academic is regarded as never having left i to go to j.

About 10,000 academics in the database have the same name. For example, there are two en-

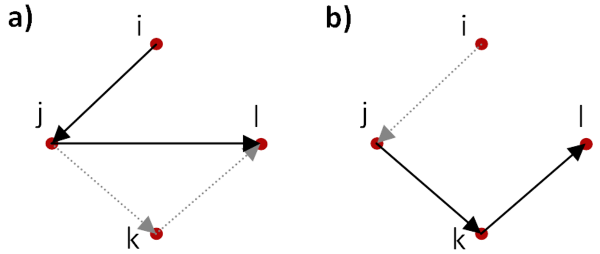

Figure 3.2: a) An academic begins at i and moves to j. The academic then proceeds to k before finishing up at l. However, there is no student at k so according to the data, the academic moves directly from j to l. b) An academic completes their PhD at i but this is not known from the MGP data. The academic goes on to have students at j, k and l, in that order. Hence, the move from i to j is missed.

tries for Lionel Mason of University of Oxford. These are almost certainly for the same person. However, there are cases where different people have the same name (e.g. Zoltán Sebestyén). Deciding which entries are genuine doubles of the same person would be a very time-consuming task. Only 348 academics who have students have the same name as another academic who also has students. Because this is a small number compared to the total number of academics who have students (37,000), the possible errors due to duplicates will be negligible.

Of the extracted movements, there are 26,450 in total by 20,732 academics between 1964 universities. By the nature of the MGP data, all the academics will have ties to a mathematical field. The first move is in the year 1406 (Heinrich von Langenstein from Université de Paris to Universität Wien). The last two are in 2012 which I manually discarded so the last recorded move is in 2009. I extracted the MGP data from its original SQL format into a text file and carried out all further manipulations by writing programs in C.

4 Analysis of the Movements of Academics

4.1 Random Model of Movements

Following Section 2.1.3, we can create a random graph model with the same degree sequence as the movement network and no self-edges. Self-edges are disallowed since an academic must move from one university to a different university.

Denote the number of moves from i to j in the data by μ_{ij} such that the adjacency matrix for the movement network is μ. The degree sequence constraints (2.7) in the model are then given by

$$O_i = \sum_j \mu_{ij} \ , \ I_i = \sum_j \mu_{ji} \qquad (4.1)$$

From (2.9), the most probable adjacency matrices in the model are the ones for which $A_{ij} \in \{0, 1\}$ for all i and j (and have no self-edges). However, these may not be possible given the degree sequence constraints (4.1). To find a most probable A possible, I developed Algorithm 1.

Algorithm 1

1) Start by taking $A = \mu$, define a variable *counter* and set *counter* $= 0$.

2) Choose the largest A_{ij} and then select two other universities, k and l, to maximise $\delta = (A_{ij} + A_{kl}) - (A_{kj} + A_{il})$.

3) Exchange weights (see Fig. 4.1) and increment *counter*:

$counter \to counter + 1$

- If counter $\leq$ 3000 and $A_{kl} > \delta/4$, round $\delta/4$ down to the nearest integer,

$$A_{ij} \to A_{ij} - \delta/4, \quad A_{kl} \to A_{kl} - \delta/4$$
$$A_{il} \to A_{il} + \delta/4, \quad A_{kj} \to A_{kj} + \delta/4$$

- Else if counter ≤ 3000 and $A_{kl} \leq \delta/4$,

$$A_{ij} \to A_{ij} - A_{kl} + 1, \quad A_{kl} \to 1$$
$$A_{il} \to A_{il} + A_{kl} - 1, \quad A_{kj} \to A_{kj} + A_{kl} - 1$$

- Else (counter > 3000)

$$A_{ij} \to A_{ij} - 1, \quad A_{kl} \to A_{kl} - 1$$
$$A_{il} \to A_{il} + 1, \quad A_{kj} \to A_{kj} + 1$$

4) Display *counter* and largest A_{ij}. If largest $A_{ij} > 1$ return to 2), else stop.

Prior to running Algorithm 1 on the movement network, there was no guarantee it would stop; trial and error was used in my development of the algorithm. The algorithm constructs a most probable adjacency matrix when *counter* reaches

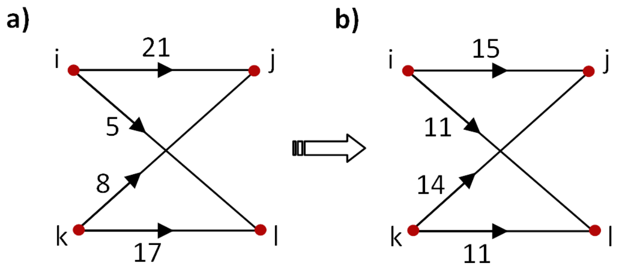

Figure 4.1: The edge weights between four nodes a) before and b) after one loop of Algorithm 1 (for counter < 3000). Here, $\delta = 25$ so i and k are decreased by 6 and j and k are increased by 6.

a value of 7401. The matrix found by the algorithm will be labelled $\widetilde{\mu}$.

The clustering coefficient, C, (2.4) is 0.255 for μ and 0.259 for $\widetilde{\mu}$, which are very similar. However, the mean local clustering coefficient, C_L, (2.5) is 0.321 for μ and 0.652 for $\widetilde{\mu}$. These results might seem surprising because real networks often have higher clustering coefficients than they would by chance [5] but this is a consequence of the random model preferring single edges between universities: nodes connect to as many other nodes as possible. In the network generated by Algorithm 1 with adjacency matrix $\widetilde{\mu}$, C_L is more than double the clustering coefficient, C. To understand this, note the high-degree nodes will connect to many other nodes. These will necessarily include low-degree nodes since there are a limited number of high-degree nodes. The low-degree nodes will predominantly connect to high-degree nodes and thus tend to have high local clustering coefficients. The low-degree nodes will not be connected to many other nodes so the high-degree nodes will have a lower local clustering coefficient. There are far more low-degree than high-degree nodes and their contribution will dominate the mean local clustering coefficient, C_L.

As well as a difference in the mean local clustering coefficient between μ and $\widetilde{\mu}$, using (2.9), I found that μ is 10^{6430} times less probable than $\widetilde{\mu}$ in the configuration model considered here. The probability of μ in the model must therefore be less than 10^{-6430}: it is very unlikely that this is a suitable model for the movement network. Nevertheless, results from this model are useful in Section 4.2.

4.2 The Influence of Supervisors

Those who have read "Surely you're joking Mr. Feynman!" [17] will recall Professor Slater advising Feynman,

"... you should go to some other university. You should find out how the rest of the world is."

Feynman later concludes,

So MIT was good, but Slater was right to warn me to go to another university for my graduate work. And I often advise my students the same way.

This is an example of a supervisor's view being passed down to their students and then on to their students' students - as one might expect. There are, however, differences in culture between countries. Academics in the US are commonly encouraged to change institutions between different stages of their careers [16] while in the UK this is much less the often case. Using the data available, the influence a supervisor has on their students can be estimated.

An academic's *supervisor* is the academic who supervised their PhD. An academic who has supervised at least one PhD at a university other than the one at which they completed their PhD is *mobile;* an academic who has not is *static.* To begin with, consider the number of academics with mobile supervisors N_m, of which M_m are mobile, and the number of academics with static supervisors N_s, of which M_s are mobile. Academics who, according to the MGP data, have not supervised a PhD are not included in the data used in this section. For those that are considered,

$$\frac{M_m}{N_m} = \frac{15042}{18196} \approx 0.827 \tag{4.2}$$

$$\frac{M_s}{N_s} = \frac{3083}{4608} \approx 0.669 \tag{4.3}$$

$$\Delta := \frac{M_m}{N_m} - \frac{M_s}{N_s} \approx 0.158 \tag{4.4}$$

In the sample considered here, the above calculation shows that an academic with a mobile supervisor is about 16% more likely to be mobile than an academic with a static supervisor. Note that for an academic with two supervisors, if one supervisor is mobile and the other is static, I regard the academic as having a mobile supervisor.

The technique set out in Section 2.2 can be used to find if, generally, an academic with a mobile supervisor is more likely to be mobile. I perform two calculations: I test if it is appropriate to assign a different probability of being mobile to academics with mobile supervisors, p_m, than to academics with static supervisors, p_s, such that $p_s \neq p_m$ (opposed to both groups having the same probability, $p_s = p_m = p$). Then, assuming $p_s \neq p_m$, I test the likelihood of $p_s < p_m$.

The ratios of interest are then given by

$$\frac{pr(p_s \neq p_m|d)}{pr(p_s = p_m|d)} \quad \text{and} \quad \frac{pr(p_s < p_m|d)}{pr(p_s > p_m|d)} \tag{4.5}$$

To calculate the first ratio of (4.5), the top and bottom terms are expressed as integrals, analogous to (2.17):

$$pr(p_s \neq p_m|d) = \int_0^1 d\widetilde{p}_m \int_0^1 d\widetilde{p}_s \; pr(p_m = \widetilde{p}_m, p_s = \widetilde{p}_s|d) \tag{4.6}$$

$$pr(p_s = p_m|d) = \int_0^1 dp \; pr(p_m = p, p_s = p|d) \tag{4.7}$$

The right-hand side terms of these equations are dealt with using Bayes' theorem (2.19) and the procedure laid out from there in Section 2.2 is followed down to (2.23).

If the movements being considered are assumed to be a random sample, the likelihood function, $pr(d|p_m, p_s)$, could be taken to be analogous to the situation of flipping two different coins: a binomial distribution for academics with mobile supervisors and another for academics with static supervisors. In this case, the relevant data, d, is given by the numbers N_m, M_m, N_s, M_s. Make the definitions $q_i := 1 - p_i$ and $S_i := N_i - M_i$ such that

$$pr(d|p_m = \widetilde{p}_m, p_s = \widetilde{p}_s) = \gamma \, \widetilde{p}_m^{\,M_m} \widetilde{q}_m^{\,S_m} \widetilde{p}_s^{\,M_s} \widetilde{q}_s^{\,S_s} \tag{4.8}$$

where γ is a normalisation constant. Putting this together, the ratio of (4.5) can be expressed as

$$\frac{pr(p_s \neq p_m|d)}{pr(p_s = p_m|d)} = \frac{\int_0^1 d\widetilde{p}_m \left[\widetilde{p}_m^{\,M_m} \widetilde{q}_m^{\,S_m}\right] \int_0^1 d\widetilde{p}_s \left[\widetilde{p}_s^{\,M_s} \widetilde{q}_s^{\,S_s}\right]}{\int_0^1 dp \left[p^{(M_m + M_s)} q^{(S_m + S_s)}\right]} \tag{4.9}$$

where the γ and $pr(d)$ terms have cancelled.

I evaluated the integrals numerically using the trapezium rule. Assuming the moves extracted from the MGP data are correct and that the likelihood function (4.8) is appropriate to use, the probability of an academic's mobility to be correlated with their supervisor's ($p_s \neq p_m$) is 10^{109} times more likely than not. Assuming then that $p_s \neq p_m$, the correlation is 10^{113} time more likely to be positive than negative.

To account for errors in the data, I altered the numbers N_m, M_m, N_s, M_s entered into the analysis. Those who are mobile according to the data will almost certainly have moved at least once. Those who are static according to the data might actually have moved - perhaps data that would indicate a move is not available. I found that if 600 static academics with static supervisors actually have mobile supervisors, Δ would be reduced to 3.1%. However, it would still be over 100 times more likely for an academic's mobility to be correlated with their supervisor's than not, and over 100 times more likely to be positively so than negatively so.

The above analysis is based on (4.8), a likelihood function appropriate for random samples. The PhDs recorded in the MGP data are mostly since 1930 and predominantly from reputable US universities. Additionally, academics who are currently in the early stages of their careers may become mobile in the future. It is unlikely then that the likelihood function used is a good choice.

Whether there is correlation between an academic's mobility and their supervisor's can be asked in the case of more specific groups of academics for whom the likelihood function (4.8) might be more appropriate. Four groups are considered in Table 1. All agree on positive correlation between an academic's and their supervisor's mobilities if correlation is present. The first two groups, grouped by year, conclude that there is correlation while the last two groups, grouped by universities, give no conclusive indication of whether there is correlation.

Group	Δ	$\frac{pr(p_s \neq p_m\|d)}{pr(p_s = p_m\|d)}$	$\frac{pr(p_s < p_m\|d)}{pr(p_s > p_m\|d)}$
1945-1959	0.113	$> 10^4$	$> 10^6$
1700-1929	0.201	$> 10^7$	$> 10^9$
Oxf. or Camb.	0.084	$< 10^1$	$> 10^2$
Pr., St. or Ha.	0.048	$< 10^1$	$> 10^2$

Table 1: Four groups of academics are tested. The groups 1945-1959 and 1700-1929 contain academics who completed their PhD theses in those periods. The third and fourth groups are made up of academics who completed their PhD at Oxford or Cambridge (UK) and Princeton, Stanford or Harvard (US) respectively.

For the first two groups, the academics will have retired from their careers or at least be nearing retirement such that they and their supervisors are unlikely to change their state of mobility. For the last two groups, a greater number of future changes in mobilities of the academics and their supervisors may be expected. I excluded academics who completed their PhDs after 1990 from the last two groups and found no stronger evidence of correlation.

Suppose movements between universities are assortatively mixed by country (see Section 2.1.4). That is, academics tend to supervise students at universities in the same country as their own PhD university. If the mobility of an academic is dependent on the country as mentioned at the start of Section 4.2, a positive correlation would be found in groups of academics from several countries but not necessarily in groups containing academics from only one country. The results in Table 1 can be seen to be consistent with assortative mixing by country in the movement network.

The modularity (2.12) is a measure of whether significant assortative mixing by country is present. For the configuration model without self-edges considered in Section 4.1,

$$\sum_i \frac{O_i I_i}{T} \approx 120\ ,\ T = 26450 \qquad (4.10)$$

so (2.11) is a good approximation. Using this, I found $Q = 0.444$ for the movement network. Using the same approximation for the network generated by Algorithm 1 – which would be expected to have a zero modularity – I found $Q = 0.101$. This modularity is smaller than that for the movement network, consistent with the approximation being reasonable and, hence, there being greater assortative mixing by country in the movement network than would be expected by chance.

4.3 Movement Between US Universities

I considered a network of movements only between 231 US universities, the US universities that have at least one move to or from them, see Fig. 4.2.

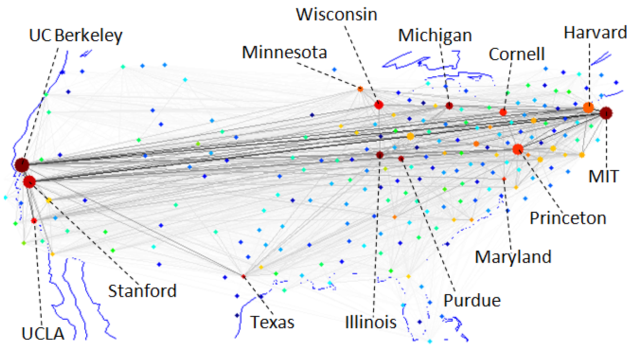

Figure 4.2: A geographically inspired, undirected graph of the movements between 231 US universities. This uses the code of S. Myers [6]. The darkness of the edges indicates the total number of moves between universities (dark for many, light for few). Universities are coloured according to their in-degree (red for high, blue for low). The universities are sized according to their out-degree (large for high, small for low). Major universities are labelled. See Appendix A for the full names.

For the network of movements between 231 US universities, I developed a model in which an academic travels around the network in discrete steps. At each step, the academic moves from their current university to a different university. The probability of the academic going from i to j at any step is defined as

$$P_{ij} := \frac{\mu_{ij}}{\sum_j \mu_{ij}} \qquad (4.11)$$

Write the probability of the academic being at university i at step t is written as W_i^t. The probability of the academic to arrive at j from i at step $t+1$ is then given by $W_i^t P_{ij}$. Generally, we have a Markov chain [18]:

$$W_j^{t+1} = \sum_i W_i^t P_{ij} \qquad (4.12)$$

Writing W_i^t as a row vector and P_{ij} as a matrix

$$W^{t+n} = W^t (P)^n \qquad (4.13)$$

As $t \to \infty$, we might expect this to reach equilibrium:

$$W^\infty = W^\infty P \qquad (4.14)$$

Formally, this requires that P has a left eigenvector [19] with eigenvalue equal to one. To get sensible results, universities for which there were zero moves to or from had to be discarded, leaving 187 universities (for names, see Appendix B). For these universities, I numerically found that P has only one left eigenvector with eigenvalue equal to one – there is a unique W^∞. My numerical calculation

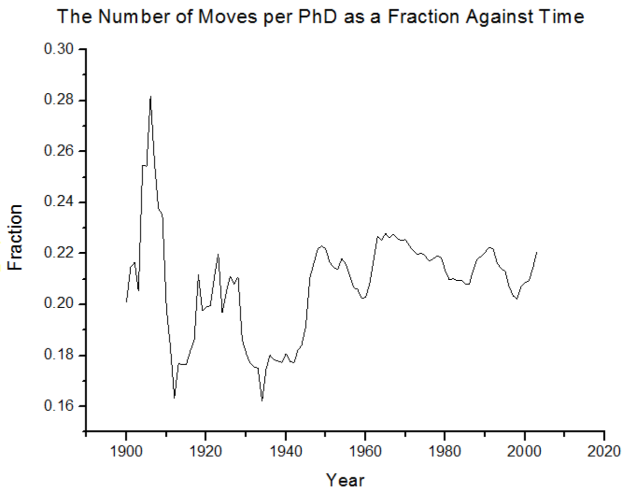

Figure 4.3: The 6-year moving total of movements between all universities as a fraction of the 6-year moving total of PhDs used to extract them.

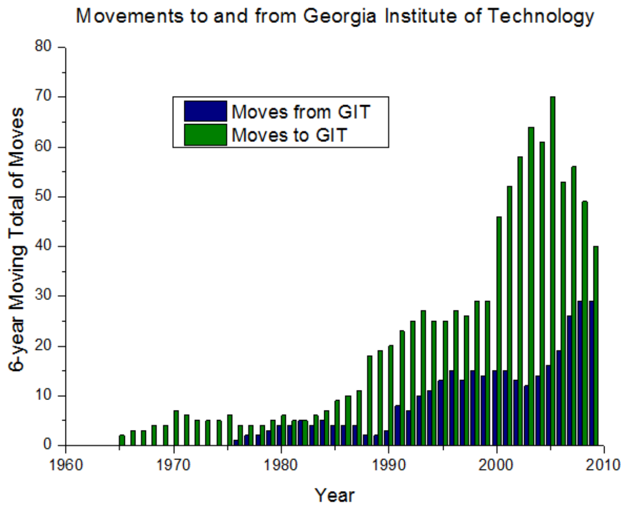

Figure 4.4: Georgia Institute of Technology

of W^∞ found that $W_i^\infty > 0$ for all i (for values, see Appendix B).

Since W^∞ is unique and all elements are positive, it is not possible to divide the nodes up into two groups such that there are edges only in one direction between the groups. Equivalently, there must be at least one cycle containing each pair of nodes. This means that the 187 US universities do not divide into groups such that academics at universities in one group cannot move to universities in another group.

4.4 Time-dependent Movement Trends

The movements extracted from the MGP data are dated by year. As discussed in Section 3.2, the year is an upper bound on the date of the move. With this in mind, the movements can be used to study past events.

In Fig 4.3, the 6-year moving total of moves is given as a fraction of the 6-year moving total of PhDs recorded in MGP. The 6-year moving total of moves or PhDs in year y is the sum of the moves or PhDs completed in each year from $y-5$ to y inclusive. Only PhDs registered in MGP that include a year, a university and at least one supervisor are counted.

The fraction has been stable, with a value of 0.213±7%, since the end of the Second World War (1946–2003) and during 1918–1928. The fraction displays dips of about 17% in the periods 1911–1917 and 1935–1945, which are the periods leading up to and during the First and Second World Wars respectively. Given the uncertainty in the year of the movements, any inference from Fig.4.3 should be made with care.

The large fluctuations between 1900 and 1930 will be partly attributable to the small numbers of moves being considered. Due to the method used to extract movements from the PhD data, nothing accurate can be inferred about movements after 2003: an apparent movement by an academic in 2004 would be invalidated if the academic returned to their previous university in 2009 and the volume of MGP data available after 2008 sharply falls.

During the mid-80s, the activities of Georgia Institute of Technology's maths department expanded from services to research with the hiring of Jack Hale as head of department [16][1]. Consistent with the expansion, in Fig. 4.4, the number of moves to Georgia Institute of Technology by can be seen to escalate in the late 80s to early 90s.

The Academy of Sciences of the GDR was known as The German Academy of Sciences in Berlin until 1972 and was closed in 1991 [20]. The movement data show that all movements to The Academy of Sciences of the GDR took place between 1973 and 1991 (see Fig. 4.5), which is consistent with the name change.

In many cases, the trend followed by the *moves from* a university is similar to the trend followed by the *moves to* a university but with a lag of 5-20 years. Such a lag can be seen in both Fig. 4.4

[1] M. A. Porter was a member of staff at Georgia Institute of Technology.

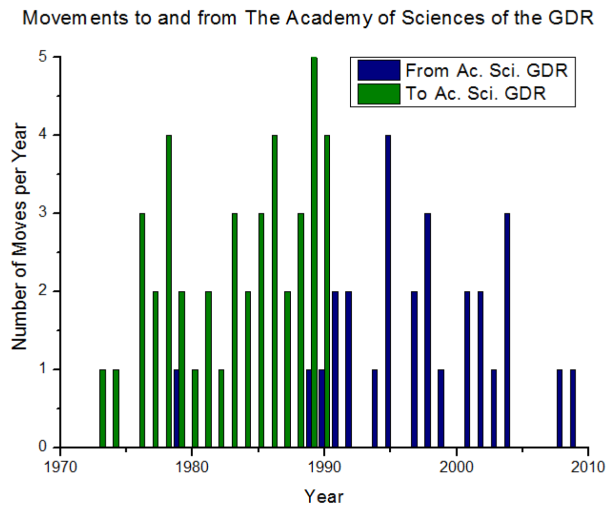

Figure 4.5: The Academy of Sciences of the GDR

and Fig. 4.5. This makes sense given the time required to leave one university and supervise a PhD at another.

5 Conclusions and Discussion

In this report, I have studied the movements of academics between universities that I extracted from the Mathematics Genealogy Project (MGP) data.

I have shown a configuration model, without self-edges, of the movements is unable to support several features of the movement network: The mean local clustering coefficient, C_L, and modularity by country, Q, are respectively 0.321 and 0.444 for the movement network but 0.652 and 0.101 for a most probable network of the model. The probability of the MGP movements occurring in the model is negligible. Hence, I have demonstrated that some factors such as job vacancies, reputation, geographical location, supervisor, etc. must have a significant influence on the large-scale movements.

In the MGP movements, I found that academics with mobile supervisors are 16% more likely to be mobile than those with static supervisors. I applied Bayesian hypothesis testing methods and, assuming the movements could be treated as a random sample, found conclusively that a supervisor's mobility would be positively correlated with their students. To check the robustness of the random sample assumption, I considered four sub-samples and showed that for two there was conclusive positive correlation but, for the other two, the correlation was inconclusive. Following this, I have suggested that assortative mixing could be the most significant cause of correlation.

Using a Markov chain, I have shown that there is a cycle containing each pair of nodes in the network of movements between 187 US universities considered in Section 4.3. The implication of this result is that universities do not form closed groups between which there is no exchange of academics, suggesting a good exchange of academic culture.

I found that the ratio of moves to completed PhDs has remained at $1 : 4.7 \pm 0.3$ since the Second World War, during which the ratio was higher than $1 : 5.5$. I have shown how temporal trends in the movements of academics are consistent with the development of Georgia Institute of Technology's maths department in the 1980s and the name changes of The Academy of Sciences of the GDR in the second half of the 20th century.

It would be insightful to find whether the likelihood function for an academic's mobility (4.8) should depend on the country in which an academic's PhD was completed. If it did, the dependence would give some indication whether assortative mixing by country is the most significant cause of correlation between an academic's mobility and their supervisor's.

A natural progression from studying mobility would be to address whether exceptional mathematicians have exceptional students. Testing for correlation between academics and their supervisors being awarded a prestigious mathematics prize [21] would be one way to proceed. If the gender of an academic could be discerned from their name, the data would be of use in equal opportunities studies on academia.

With future MGP data, movements could be used to assess the impact of funding cuts to universities in the UK or elsewhere by examining the departure of highly-regarded academics.

Acknowledgments

I thank my supervisor, Mason Porter, and his colleague, Peter Mucha for their useful suggestions throughout the project; Mitch Keller for use of the MGP data; Sean Myers for his geographically inspired network code; Devinder Sivia for his advice on Bayesian data analysis techniques; and Felix Flicker for helpful tips on typesetting.

References

[1] Paton G. *Oxford Beaten by UCL in League Table* (The Telegraph, 2009) http://www.telegraph.co.uk/education/6269348/Oxford-beaten-by-UCL-in-league-table.html (accessed March 2011)

[2] Bowman N. & Bastedo M. *Getting on the Front Page: Organizational Reputation, Status Signals, and the Impact of U.S. News and World Report Rankings on Student Decisions* http://www-personal.umich.edu/~bastedo/papers/BowmanBastedo.ResHE2009.pdf (accessed March 2011)

[3] The Mathematics Genealogy Project http://www.genealogy.ams.org (accessed March 2011)

[4] Malmgren R.D., Ottino J.M. & Amaral L.A.N. *The Role of Mentorship in Protégé Performance* (Nature, Vol. 465, pages 622–626, 2010)

[5] Newman M.E.J. *Networks: An Introduction* (Oxford University Press, 2010)

[6] Myers S., Mucha P.J. & Porter M.A. *Mathematical Genealogy & Department Prestige*, to appear in Chaos (Gallery of Nonlinear Images, 2011)

[7] Wasserman S. & Faust K. *Social Network Analysis: Methods and Applications* (Cambridge University Press, 1994)

[8] Hartmann A.K. & Weigt M. *Phase Transitions in Combinatorial Optimization Problems* (Wiley-VCH, 2005)

[9] Newman M.E.J. *The physics of networks* (Physics Today, Nov. 2008, page 33)

[10] Strogatz S.H. *Exploring complex networks* (Nature, Vol. 410, pages 268-276, 2001)

[11] Sivia D.S., Skilling J. *Data Analysis A Bayesian Tutorial* (Oxford University Press, 2008)

[12] Newman M.E.J. *Assortative mixing in networks* (2002) http://arxiv.org/PS_cache/cond-mat/pdf/0205/0205405v1.pdf (accessed March 2011)

[13] Newman M.E.J. *Modularity and community structure in networks* (2006) http://arxiv.org/PS_cache/physics/pdf/0602/0602124v1.pdf (accessed March 2011)

[14] Jeffreys H. *The Theory of Probability* (Oxford University Press, 1939)

[15] Jackson A. *A Labor of Love: The Mathematics Genealogy Project* (Notices of the AMS, Vol. 54, No. 8, 1002-1003, 2007) http://www.ams.org/notices/200708/tx070801002p.pdf (accessed April 2011)

[16] Private correspondence with M. A. Porter

[17] Feynman R.P. *"Surely you're joking Mr. Feynman!" Adventures of a curious character* (Vintage, 1992)

[18] Reichl L. E. *A Modern Course in Statistical Physics* (Wiley-VCH, 2009)

[19] Weisstein, Eric W. *Left Eigenvector* (MathWorld–A Wolfram Web Resource) http://mathworld.wolfram.com/LeftEigenvector.html (accessed March 2011)

[20] Academy of Sciences of the GDR http://www.math.dartmouth.edu/~euler/historical/places/berlin.html (accessed March 2011)

[21] Weisstein, E.W. *Mathematics Prizes* (MathWorld–A Wolfram Web Resource) http://mathworld.wolfram.com/MathematicsPrizes.html (accessed March 2011)

Appendix

A Universities in Figure 4.2

West to East (top):

University of California, Berkeley
University of Minnesota-Minneapolis
University of Wisconsin-Madison
University of Michigan
Cornell University
Harvard University

West to East (bottom):

University of California, Los Angeles
Stanford University
University of Texas at Austin
University of Illinois at Urbana-Champaign
Purdue University
University of Maryland at College Park
Princeton University
Massachusetts Institute of Technology

B 187 US universities

Ordered by net movement ranking

i	$\mathbf{W}_i$	University
1	2.28×10^{-2}	North Carolina State Uxniversity
2	2.00×10^{-2}	University of Texas at Austin
3	2.00×10^{-2}	Purdue University
4	1.80×10^{-2}	Texas A&M University
5	1.62×10^{-2}	University of Illinois at Urbana-Champaign
6	1.56×10^{-2}	Georgia Institute of Technology
7	1.53×10^{-2}	University of Maryland, College Park
8	1.48×10^{-2}	University of Michigan
9	1.43×10^{-2}	Virginia Polytechnic Institute and State University
10	1.30×10^{-2}	University of Florida
11	1.30×10^{-2}	The Pennsylvania State University
12	1.27×10^{-2}	Colorado State University
13	1.23×10^{-2}	George Mason University
14	1.23×10^{-2}	Rutgers University, New Brunswick
15	1.21×10^{-2}	University of South Florida
16	1.21×10^{-2}	The Johns Hopkins University
17	1.20×10^{-2}	University of Wisconsin-Madison
18	1.17×10^{-2}	Auburn University
19	1.16×10^{-2}	Iowa State University
20	1.14×10^{-2}	The Florida State University
21	1.10×10^{-2}	Arizona State University
22	1.09×10^{-2}	University of Iowa
23	1.09×10^{-2}	The University of North Carolina at Chapel Hill
24	1.06×10^{-2}	Princeton University
25	1.03×10^{-2}	University of Minnesota-Minneapolis
26	1.02×10^{-2}	Temple University
27	9.51×10^{-3}	Montana State University
28	9.41×10^{-3}	University of Colorado at Boulder
29	9.36×10^{-3}	University of California, Berkeley
30	9.34×10^{-3}	The Ohio State University
31	8.92×10^{-3}	University of Arizona
32	8.73×10^{-3}	University of South Carolina
33	8.64×10^{-3}	Oregon State University
34	8.59×10^{-3}	University of Louisiana at Lafayette
35	8.57×10^{-3}	University of California, Los Angeles
36	8.45×10^{-3}	Cornell University
37	8.35×10^{-3}	Texas Tech University
38	8.22×10^{-3}	University of Houston
39	8.16×10^{-3}	Massachusetts Institute of Technology
40	8.13×10^{-3}	University of Pittsburgh
41	8.08×10^{-3}	Michigan State University
42	8.05×10^{-3}	Stanford University
43	7.96×10^{-3}	Vanderbilt University
44	7.93×10^{-3}	University of Missouri - Columbia
45	7.80×10^{-3}	University of Washington
46	7.78×10^{-3}	The Louisiana State University
47	7.64×10^{-3}	Clemson University
48	7.60×10^{-3}	University of North Texas
49	7.59×10^{-3}	Carnegie Mellon University
50	7.56×10^{-3}	Syracuse University
51	7.55×10^{-3}	University of California, Riverside
52	7.14×10^{-3}	Oklahoma State University
53	7.01×10^{-3}	University of Alabama-Tuscaloosa
54	6.97×10^{-3}	University of Wisconsin-Milwaukee
55	6.94×10^{-3}	University of Georgia

56	6.90×10^{-3}	Kansas State University
57	6.85×10^{-3}	State University of New York at Stony Brook
58	6.83×10^{-3}	State University of New York at Buffalo
59	6.71×10^{-3}	Harvard University
60	6.68×10^{-3}	University of Delaware
61	6.64×10^{-3}	The University of Connecticut
62	6.61×10^{-3}	Columbia University
63	6.56×10^{-3}	Baylor University
64	6.52×10^{-3}	Missouri University of Science & Technology
65	6.41×10^{-3}	University of Central Florida
66	6.34×10^{-3}	University of Nebraska-Lincoln
67	6.33×10^{-3}	Southern Methodist University
68	6.32×10^{-3}	University of Kentucky
69	6.24×10^{-3}	University of Massachusetts Amherst
70	6.03×10^{-3}	University of Tennessee - Knoxville
71	5.97×10^{-3}	University of Pennsylvania
72	5.59×10^{-3}	University of California, San Diego
73	5.46×10^{-3}	New York University
74	5.39×10^{-3}	Northwestern University
75	5.37×10^{-3}	University of Utah
76	5.32×10^{-3}	Emory University
77	5.31×10^{-3}	The University of Oklahoma
78	5.30×10^{-3}	Wayne State University
79	5.29×10^{-3}	Rensselaer Polytechnic Institute
80	5.23×10^{-3}	University of Texas at Arlington
81	5.16×10^{-3}	Tulane University
82	5.10×10^{-3}	Indiana University
83	5.07×10^{-3}	New Mexico State University
84	4.96×10^{-3}	University of California, Davis
85	4.83×10^{-3}	University of New Mexico
86	4.74×10^{-3}	University of Montana
87	4.73×10^{-3}	The College of William and Mary
88	4.68×10^{-3}	Claremont Graduate University
89	4.56×10^{-3}	Washington State University
90	4.51×10^{-3}	Rice University
91	4.51×10^{-3}	The University of Texas at Dallas
92	4.50×10^{-3}	University of California, Santa Barbara
93	4.49×10^{-3}	University of Virginia
94	4.46×10^{-3}	The University of Chicago
95	4.32×10^{-3}	Kent State University
96	4.30×10^{-3}	Brown University
97	4.20×10^{-3}	The University of Memphis
98	4.12×10^{-3}	Duke University
99	3.98×10^{-3}	Dartmouth College
100	3.96×10^{-3}	Case Western Reserve University
101	3.94×10^{-3}	University of Arkansas
102	3.92×10^{-3}	The George Washington University
103	3.89×10^{-3}	City University of New York
104	3.86×10^{-3}	Florida Institute of Technology
105	3.86×10^{-3}	Boston University
106	3.81×10^{-3}	University of Oregon
107	3.81×10^{-3}	Bowling Green State University
108	3.79×10^{-3}	University of Miami
109	3.67×10^{-3}	West Virginia University
110	3.64×10^{-3}	Western Michigan University
111	3.61×10^{-3}	University of Colorado at Denver
112	3.58×10^{-3}	University of Southern California
113	3.47×10^{-3}	University of Alabama-Birmingham
114	3.46×10^{-3}	Yale University
115	3.35×10^{-3}	University of Wyoming
116	3.30×10^{-3}	Lehigh University
117	3.29×10^{-3}	University of California, Irvine
118	3.27×10^{-3}	University of Kansas
119	3.26×10^{-3}	California Institute of Technology
120	3.22×10^{-3}	Northern Illinois University
121	3.17×10^{-3}	The University of Rochester
122	3.08×10^{-3}	University of Illinois at Chicago
123	3.05×10^{-3}	University of New Hampshire
124	3.04×10^{-3}	Washington University
125	2.89×10^{-3}	Illinois Institute of Technology
126	2.86×10^{-3}	University of Cincinnati
127	2.82×10^{-3}	The American University
128	2.76×10^{-3}	University of Notre Dame
129	2.69×10^{-3}	Stevens Institute of Technology
130	2.65×10^{-3}	Air Force Institute of Technology
131	2.50×10^{-3}	Southern Illinois University at Carbondale
132	2.41×10^{-3}	Northeastern University
133	2.35×10^{-3}	Clarkson University
134	2.35×10^{-3}	University of Maryland, Baltimore County
135	2.33×10^{-3}	North Dakota State University
136	2.32×10^{-3}	University of Idaho

137	2.28×10^{-3}	University of Hawaii
138	2.13×10^{-3}	University of California, Santa Cruz
139	2.11×10^{-3}	Old Dominion University
140	2.08×10^{-3}	State University of New York at Binghamton
141	2.07×10^{-3}	State University of New York at Albany
142	2.01×10^{-3}	Virginia Commonwealth University
143	1.95×10^{-3}	Wichita State University
144	1.74×10^{-3}	University of Alabama in Huntsville
145	1.68×10^{-3}	Louisiana Tech University
146	1.67×10^{-3}	St. Louis University
147	1.64×10^{-3}	Naval Postgraduate School
148	1.62×10^{-3}	Drexel University
149	1.59×10^{-3}	Polytechnic University
150	1.54×10^{-3}	New Jersey Institute of Technology
151	1.26×10^{-3}	Illinois State University
152	1.24×10^{-3}	Utah State University
153	1.18×10^{-3}	Brandeis University
154	1.17×10^{-3}	University of Rhode Island
155	1.13×10^{-3}	Tufts University
156	1.11×10^{-3}	Howard University
157	1.10×10^{-3}	Clark University
158	1.09×10^{-3}	Georgia State University
159	1.05×10^{-3}	Texas Christian University
160	1.04×10^{-3}	Worcester Polytechnic Institute
161	9.91×10^{-4}	Portland State University
162	9.76×10^{-4}	Claremont Graduate University
163	9.28×10^{-4}	University of Denver
164	9.22×10^{-4}	Wesleyan University
165	8.95×10^{-4}	University of Toledo
166	8.64×10^{-4}	New Mexico School of Mining and Technology
167	8.58×10^{-4}	Oakland University
168	8.04×10^{-4}	University of Southern Mississippi
169	7.21×10^{-4}	Brigham Young University
170	6.60×10^{-4}	Adelphi University
171	5.78×10^{-4}	University of Texas at Houston
172	5.14×10^{-4}	Yeshiva University
173	4.74×10^{-4}	Memphis State University
174	4.52×10^{-4}	Bryn Mawr College
175	4.17×10^{-4}	Rutgers University, Newark
176	2.82×10^{-4}	University of Texas at El Paso
177	2.70×10^{-4}	The Catholic University of America
178	2.36×10^{-4}	University of Colorado Health Sciences Center
179	2.36×10^{-4}	University of Texas at San Antonio
180	2.04×10^{-4}	Polytechnic Institute of New York
181	1.66×10^{-4}	The Rockefeller University
182	1.41×10^{-4}	Medical University of South Carolina
183	1.29×10^{-4}	Oregon Graduate Institute of Science & Technology
184	1.06×10^{-4}	Georgetown University
185	9.59×10^{-5}	Case Institute of Technology
186	6.02×10^{-5}	University of California, San Francisco
187	4.12×10^{-5}	Union College