AMSI VACATION RESEARCH SCHOLARSHIPS 2018-2019

Adoption of Agricultural Techniques in African Countries

Jackson Ryder

Supervised by **Professor Diane Donovan**

University of Queensland

Vacation Research Scholarships are funded jointly by the Department of Education and Training and the Australian Mathematical Sciences Institute.

AMSI Australian Mathematical Sciences Institute

Australian Government
Department of Education and Training

1 Abstract

Using data obtained from the SIMLESA program, which introduced agricultural technologies to farms in Africa, we construct a network to encode the information about the adoption of these technologies, and then use techniques from network theory to examine the structure of this network.

2 Introduction

Since 2010, the Sustainable Intensication of Maize-Legume Cropping Systems for Food Security in Eastern and Southern Africa (SIMLESA) program has been introducing sustainable agricultural practices to farmers in eastern and southern African countries, in order to increase food production CIMMYT [2010]. This is done through demonstration sites, where farmers can see the impact that the practices have on crop yield, and learn how they can implement new practices on their own farms. During this time, data was collected relating to the adoption of 7 agricultural techniques over 8 years. The aim of this paper is to determine the effectiveness of demonstration sites in relation to the adoption of agricultural practices. This will be done by creating a weight function which quantifies the similarity of adoption of pairs of farms based on this data, and then examining the structure of the network of farms by endowing it with said weighting and searching for community structure.

3 Graphs

Throughout the next two sections, definitions have been taken from Godsil and Royle [2001].

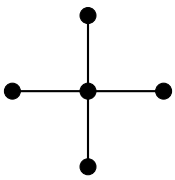
Definition 3.0.1. A **graph** $\mathcal{G}$ is a pair of sets (V, E), where elements of V are called vertices, or nodes, and $E \subseteq \{\{u, v\} | u, v \in V\}$, whose elements are referred to as edges.

An edge between vertices u and v is denoted $u \sim v$, or simply uv, and the two vertices are said to be **adjacent**. A vertex v and edge e are **incident** if $e = uv$ for some $u \in V$. An edge from a vertex to itself is called a **loop**.

A graph is said to be **simple** if it has no loops, and has at most one edge between all pairs of vertices. For a simple graph, the set of all vertices which are adjacent to u is called the **neighbourhood** of u, denoted $N(u)$, and $|N(u)|$ is called the **degree** of u, denoted as $deg(u)$.

If the elements of E are taken as ordered pairs, then the graph obtained is referred to as a **directed graph**, or **digraph**, where edge $(a, b) \neq (b, a)$.

Throughout, $\mathcal{G}$ is an undirected, simple graph, unless specified otherwise.



Visualisation of a simple graph.

The above graph is an example of a complete bipartite graph.

Definition 3.0.2. The **complete graph** K_n on n vertices is the graph where $\forall u, v \in V, uv \in E$.

Definition 3.0.3. A graph is **bipartite** if $V = X \cup Y$, where $X \cap Y = \emptyset$, and if $u \sim v$, then $u \in X, v \in Y$.

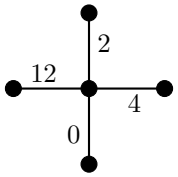
A **complete bipartite** graph $K_{m,n}$ is a bipartite graph with $|X| = m, |Y| = n$ and $\forall u \in X, v \in Y$, $u \sim v$. An important class of graphs are **stars**, which are complete bipartite graphs $K_{1,n}$. The above example is a drawing of $K_{1,4}$. In many cases it is useful to examine the distance between vertices of a graph. To define some notion of distance, we must first define a path in a graph.

Definition 3.0.4. A **path** P is an alternating sequence of vertices and edges $(v_0, e_0, v_1, e_1, ..., e_{n-1}, v_n)$, where $e_i = v_i v_{i+1}$, and no vertex is repeated.

If a path P has $v_0 = u, v_n = v$, then P is said to be a $u - v$ path.

Definition 3.0.5. A **vertex labelling** on a graph is a function $w : V \to S$, which assigns a label from S, a set of labels, to each vertex of the graph. If $S \subseteq \mathbb{R}$, then w is said to be an vertex weighting.

Edge weightings are defined in a similar manner.



An edge weighting on $K_{1,4}$.

Using paths and edge weightings, a distance measure can be imposed on the graph to define a distance between vertices.

Definition 3.0.6. The **distance** from vertex u to vertex v, $d(u, v) = \min_P\{\sum_{v_i v_{i+1} \in P} w(v_i v_{i+1})\}$, where $\min_P$ is the minimum over all $u - v$ paths and w is an edge weighting.

4 Matrices of Graphs

Information from a graph can be encoded into different matrices, where techniques from linear algebra can be used to examine the structure of the graph. Two of the most useful matrices related to a graph are the adjacency and incidence matrices.

Definition 4.0.1. The **adjacency matrix** $\mathbf{A} = [a_{ij}]$ of a graph $\mathcal{G}$ with $|V| = n$ is the $n \times n$ matrix, whose rows and columns are indexed by the vertices of $\mathcal{G}$, and

$$a_{ij} = \begin{cases} 0, & ij \notin E, \\ 1, & ij \in E. \end{cases}$$

In the case of a weighted graph, each (i, j)-entry is the value of $w(ij)$.

Definition 4.0.2. The **incidence matrix** $\mathbf{B} = [b_{ij}]$ of a graph $\mathcal{G}$ with $|V| = n$, and $|E| = e$ is the $n \times e$ matrix, whose rows and columns are indexed by the vertices and edge of $\mathcal{G}$, respectively, and

$$b_{ij} = \begin{cases} 0, & \text{if vertex i is not incident with edge j,} \\ 1, & \text{if vertex i is incident with edge j,} \\ 2, & \text{if edge j is a loop on vertex i.} \end{cases}$$

The adjacency and incidence matrices for $K_{1,4}$ are $\mathbf{A}(K_{1,4}) = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$,

and $\mathbf{B}(K_{1,4}) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

These two matrices can provide a significant amount of useful information about a graph, such as the degree of each vertex. As defined above, the **degree** of a vertex is the number of vertices it is adjacent to, and can be calculated in the following equivalent ways:

1. $deg(i) = \sum_{j=1}^{n} a_{ij}$
2. $deg(i) = \sum_{j=1}^{e} b_{ij}$

The incidence matrix contains information about the graph which relates the degrees of the vertices to the number of edges of $\mathcal{G}$.

Lemma 4.0.1. For a graph $\mathcal{G}$, with $|V| = n$ and $|E| = e$, $\sum_{i=1}^{n} deg(i) = 2e$.

Proof. To see this, consider the sum of the rows and columns of the incidence matrix $\mathbf{B} = [b_{ij}]$.
$\sum_{j=1}^{e}\sum_{i=1}^{n} b_{ij} = 2e$, as each edge is incident with two vertices, and there are e columns.
But, $\sum_{i=1}^{n}\sum_{j=1}^{e} b_{ij} = \sum_{i=1}^{n} deg(i)$, by definition 3.0.3.
These two sums are the same, and hence $\sum_{i=1}^{n} deg(i) = 2e$.

□

Corollary 4.0.1.1. Every graph has an even number of vertices of odd degree.

An important area of linear algebra is spectral theory, which can also provide useful information about a graph.

Definition 4.0.3. The **spectrum** of a graph is $\lambda_0 \leq \lambda_1 \leq ... \leq \lambda_n$ where the λ_i are the eigenvalues of $\mathbf{A}$. This ordering is valid, as a consequence of the following theorem.

Lemma 4.0.2. If $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors of a graph $\mathcal{G}$ with non-zero eigenvalues such that $\mathbf{x} \neq \mathbf{y}$, then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal.

Proof. Suppose $\mathbf{Ax} = \lambda_1\mathbf{x}$, and $\mathbf{Ay} = \lambda_2\mathbf{y}$. As $\mathcal{G}$ is undirected, $\mathbf{A}$ is symmetric. Hence $\mathbf{x}^T\mathbf{Ay} = (\mathbf{y}^T\mathbf{Ax})^T$. This implies that $\lambda_2\mathbf{x}^T\mathbf{y} = \lambda_1\mathbf{x}^T\mathbf{y}$. As $\lambda_1, \lambda_2 \neq 0, \mathbf{x}^T\mathbf{y} = 0$, and thus $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. □

Theorem 4.0.3. The eigenvalues of an undirected graph are real.

Proof. To begin, notice that the adjacency matrix of an undirected graph is real and symmetric. Suppose, for contradiction, that there is a non-real eigenvalue λ in the spectrum of $\mathcal{G}$. Then taking the complex conjugate of the equation $\mathbf{Ax} = \lambda\mathbf{x}$ tells us that $\mathbf{A}\bar{\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}}$. As eigenvalues are non-zero, $\mathbf{x}^T\bar{\mathbf{x}} > 0$. By the previous lemma, this is a contradiction, and hence $\lambda = \bar{\lambda}$. □

This theorem will prove very important when spectral methods of graph clustering are discussed later. As the adjacency matrix is indexed by the vertices of a graph $\mathcal{G}$, it can be realised as a linear transformation between maps in $\mathbb{R}^V$, the space of functions over the vertex set of $\mathcal{G}$. Hence, an eigenvector is a function whose domain is the vertices of $\mathcal{G}$, and satisfies $\mathbf{A}f = \lambda f$. Thus, $f \in \mathbb{R}^V$ is equivalent to a $|V|$-tuple, indexed by the vertices of $\mathcal{G}$, where the coordinate of the $|V|$-tuple f corresponding to vertex u is denoted $f(u)$. For vertex u and tuple $\mathbf{X}$, the u-component of $\mathbf{X}$ may also be denoted x_u.

The Laplacian Matrix is another important matrix of a graph, whose spectrum provides useful insight into the structure of a graph. To define the Laplacian, we must first define the diagonal matrix.

Definition 4.0.4. The **diagonal matrix** $\mathbf{D} = [d_{ij}$ of $\mathcal{G}$ is the $n \times n$ matrix where $d_{ij} = \delta_{ij}deg(i)$, where $\delta_{ij} = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases}$

Definition 4.0.5. The **Laplacian matrix** of a graph $\mathcal{G}$ with $|V| = n$ is the $n \times n$ matrix $\mathbf{Q} = \mathbf{D} - \mathbf{A}$.

Using the Laplacian matrix, the sum of squared difference for the components of any vector over $\mathcal{G}$ can be calculated.

Lemma 4.0.4. For an undirected graph $\mathcal{G}$ and some vector $\mathbf{X} \in \mathbb{R}^V$, $\mathbf{X}^T\mathbf{Q}\mathbf{X} = \sum_{uv \in E} (x_u - x_v)^2$.

Proof.

$$\mathbf{X}^T\mathbf{Q}\mathbf{X} = \mathbf{X}^T\mathbf{D}\mathbf{X} - \mathbf{X}^T\mathbf{A}\mathbf{X} = \sum_{i=1}^{n} x_i^2 d_{ii} - \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij}.$$

$\sum_{i=1}^{n} x_i^2 d_{ii} = \sum_{uv \in E} (x_u^2 + x_v^2)$, as each x_i^2 is counted $deg(i)$ times, and each edge uv contributes one x_u^2 and one x_v^2, where each x_u^2 is counted $deg(u)$ times, as $|N(u)| = deg(u)$.

$\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij} = 2 \sum_{uv \in E} x_u x_v$, as each $x_i x_j$ is counted twice in the double sum only if $ij \in E$.

Hence
$$\begin{aligned}\sum_{i=1}^{n} x_i^2 d_{ii} - \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij} &= \sum_{uv \in E} (x_u^2 + x_v^2) - 2 \sum_{uv \in E} x_u x_v \\ &= \sum_{uv \in E} (x_u^2 - 2x_u x_v + x_v^2) \\ &= \sum_{uv \in E} (x_u - x_v)^2.\end{aligned}$$

□

5 Communities in Networks

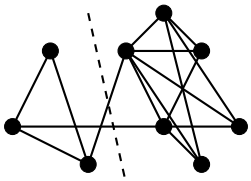
Now that we have the mathematics to construct the network it is necessary to introduce the methods with which we will examine its structure. Throughout this section, all definitions will come from Fortunato [2010] unless stated otherwise.

5.1 Methods of Community Detection

Definition 5.1.1. A **community** in a network is a partition $\mathcal{P}$ of the vertex set, where vertices in the same partite set, or cluster, are more "*similar*" to each other than to vertices from distinct clusters.

In this context, similarity is defined by some measure, such as the distance function obtained from an edge weighting. To determine the clusters there are many different methods, each with their own pros and cons. There are 4 main types of clustering procedures: graph partitioning, hierarchical clustering, partitional clustering and spectral clustering.

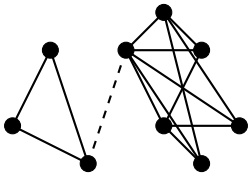
Graph partitioning relies on partitioning the vertex set into clusters such that the edges between vertices in distinct clusters is minimised. This is usually achieved through iterative bisection.



An example of a bisection of a graph

Graph bisection is an NP-hard problem, meaning there is no known algorithm to determine an optimal solution. However, there are multiple algorithms which return relatively good solutions in a reasonable amount of time. One of these is spectral bisection, which will be discussed in more detail later.

Hierarchical clustering refers to iterative processes where either edges are added and clusters merged based on a degree of high similarity, or edges are removed and clusters split based on a degree of low similarity. Hierarchical clustering methods are useful for examining multilayer community structure, where clusters may not strictly be partitions. A popular hierarchical clustering method is the algorithm proposed by Girvan and Newman [2002], where the edge centrality, a measure of the importance of an edge with respect to some network property, such as distance, is calculated for all edges of the network, and the edge with the largest centrality value is removed. This process is then repeated to determine the clusters.



An example of edge centrality with respect to vertex distance, where the dashed line represents the edge with maximum centrality.

Partitional clustering methods rely on embedding the network into a *metric* space, and maximising/minimising the value of a cost function between vertices and centroids, which are points in the space which may or may not be vertices.

A metric is a function d which satisfies the following conditions:

1. $d(x, x) = 0$,

2. $d(x, y) > 0$ if $x \neq y$,

3. $d(x, y) = d(y, x)$,

4. $d(x, z) \leq d(x, y) + d(y, z)$.

The cost function used is some function of the metric. These methods require a preassigned number of clusters, namely the number of centroids. A common example of partitional clustering is called *k-means clustering*, where, for a given initial clustering of the network, the following cost function is minimised:

$$\sum_{i=1}^{k} \sum_{v \in C_i} d(v, C_i),$$

where d is the metric for the space in which the network is embedded, and C_i is the i-th cluster. This method determines clusters which minimise the average distance between a cluster's vertices and its centroid.

Lastly, **spectral clustering** methods are similar to graph partitioning, however the eigenvalues of either the adjacency or Laplacian matrices are used in order to minimise functions of the respective matrices, based on results from spectral matrix theory. As mentioned earlier, spectral bisection is a useful method of spectral clustering.

5.2 Spectral Clustering

The procedure for spectral bisection is outlined in Donovan et al. [2008]. The goal of spectral bisection is, for a partition of V into subsets U and W such that $U \cup W = V$, and $U \cap W = \emptyset$, to minimise the *cut ratio* of the partition.

Definition 5.2.1. For a partition of V into sets U and W, the **cut ratio** $r = \frac{w(U,W)}{|U||W|}$, where $w(U,W) = \sum_{x \in U, y \in W} w(xy)$, w is a weight function and $xy \in E$.

To minimise this, the vertex set is given an arbitrary ordering $(v_1, ..., v_n)$. Then, it is recognised that if $p = \frac{|U|}{n}$, $q = \frac{|W|}{n}$ and $\mathbf{X}^T = (x_1, ...x_n)$, where $x_i = \begin{cases} q, & v_i \in U, \\ -p, & v_i \in W \end{cases}$, then the following is true as a result of lemma 3.0.4

$$r = \frac{w(U,W)}{|U||W|} = \frac{\mathbf{X}^T \mathbf{Q} \mathbf{X}}{n|\mathbf{X}|^2}.$$

This can be minimised through Lagrange multipliers, and it is shown in Donovan et al. [2008] that $r \geq \frac{\lambda}{n}$, where λ is the second smallest eigenvalue of the Laplacian. Thus, the eigenvector corresponding to λ, known as the Fielder eigenvector, can be used to minimise the cut ratio. To do so, the coordinates of the Fielder eigenvector $\mathcal{F} = (f_1, ...f_n)$ are ordered such that $f_i \leq f_{i+1}$, and the vertex set is given the same ordering. For the partite sets $U = (v_1, ..., v_k)$ and $W = (v_{k+1}, ..., v_n)$, k, referred to as the splitting index, can be chosen in a number of ways, such as to minimise the cut ratio, or to split into equally sized clusters, for example.

In terms of the adjacency matrix, this reordering creates two square blocks along the diagonal, with entries within these blocks being generally larger, and hence showing higher similarity, than entries outside of the blocks.

5.3 Dissimilarity of Clusterings

Now that we have the tools to search for community structure in a network, it is necessary to introduce methods to compare different clusterings. To evaluate the dissimilarity of two clusterings, C_1 and C_2, of a network, a measure called the Rand Index is introduced, and is as in Donovan et al. [2008]. To begin, an equivalence relation on the clusterings is defined as follows. A pair of vertices $\{u, v\}$ are said to be **consistent** if u and v are either in the same cluster for both clusterings, or are in different clusterings. If $\{u, v\}$ are called **inconsistent** if they are not consistent. Let $R^+ = \{\{u, v\} | u \text{ and } v \text{ are consistent vertices in } V\}$, and $R^- = \{\{u, v\} | u \text{ and } v \text{ are inconsistent vertices in } V\}$. The Rand Index is then defined as follows.

Definition 5.3.1. The **Rand Index** for clusterings C_1 and C_2 is defined as $R(C_1, C_2) = 1 - |R^+|/(|R^+| + |R^-|)$.

It is clear that $R \in [0, 1]$, with $R = 0$ if $C_1 = C_2$, and $R = 1$ if all pairs of vertices are inconsistent. Throughout the next two sections, the method of community detection outlined in section 4 is applied to the examine the data obtained from the SIMLESA survey in Ethiopia.

6 Construction of the Network

Using the data obtained in 2010 and 2013, each of the 1122 farms from the Ethiopian survey were treated as a vertex of the network. These vertices were labelled with a 7-tuple with binary entries, where each entry corresponds to whether or not said farm has adopted one of the seven agricultural techniques. All edges were added to the network, to create a complete graph. These edges were then given weight $7 - \sum_{i=1}^{7} |x_i - y_i|$, where $(x_1, ..., x_7)$ and $(y_1, ..., y_7)$ are the vertex labels of the vertices incident with each edge. This shows the number of technologies which both farms agree on. From this, the weighted adjacency matrix was created, and the above process of spectral clustering was used to examine the community structure produced by the edge weighting. The splitting index was chosen to be $\lfloor \frac{m}{2} \rfloor$, where m is the size of the cluster.

7 Results

7.1 Previous Results

Previously it was shown that the clusters are spatially localised, in the sense that clustering was unchanged when edges were restricted to farms within a certain threshold distance from each other. It was also shown that there was no correlation between the distance to the nearest demonstration site and increased adoption.

7.2 Champion Adopters

We now investigate the influence of farms who adopted 6 or 7 technologies, referred to as champion adopters, on their neighbours.

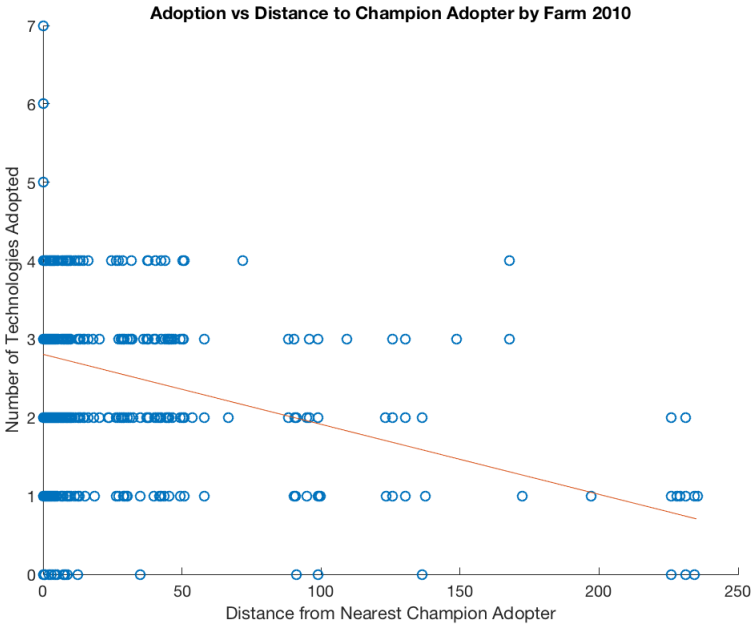


Figure 1: Distance to Champion Adopter vs. Adoption 2010

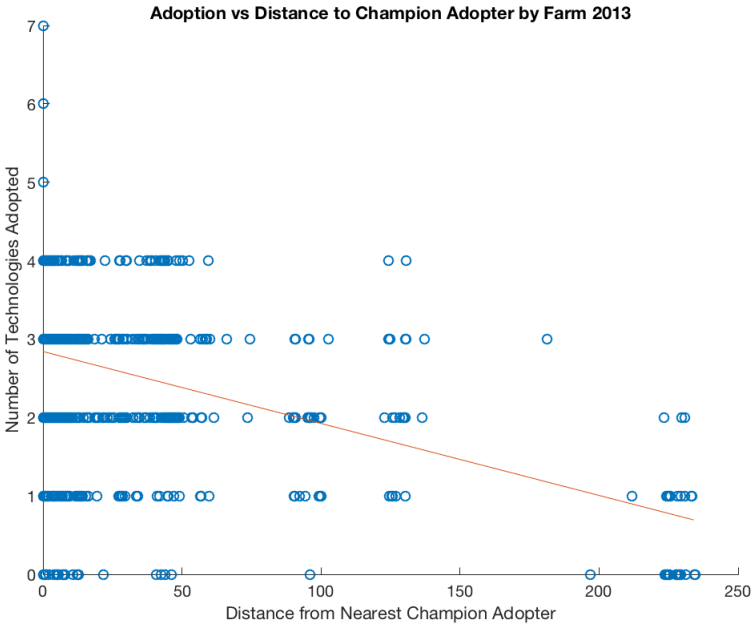


Figure 2: Distance to Champion Adopter vs. Adoption 2013

Figures 1 and 2 show the distance to the nearest champion adopter and the number of technologies adopted by each farm, as well as a line of best fit. There is no significant correlation between the two ($R^2 = 0.2490$ and $R^2 = 0.1323$, respectively), meaning that the distance to the nearest champion adopter had little effect on individual farms adoption.

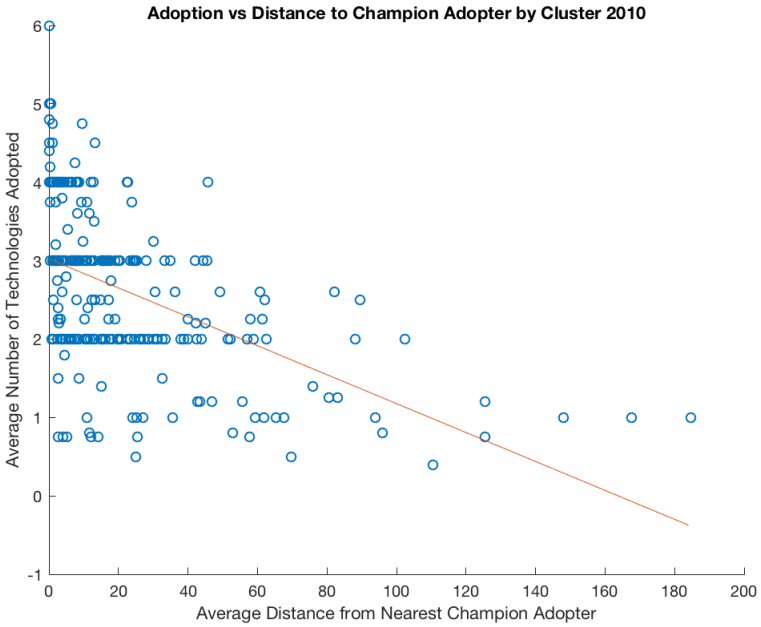


Figure 3: Average Distance to Champion Adopter vs. Adoption 2010

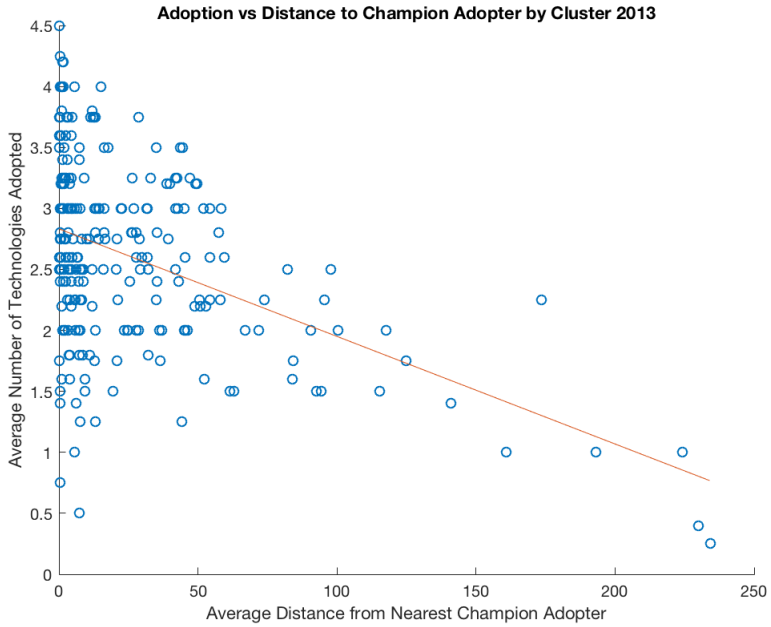


Figure 4: Average Distance to Champion Adopter vs. Adoption 2013

Figures 3 and 4 show the average distance to the nearest champion adopter and the average number of technologies adopted for each cluster. Again, there is no significant correlation between these variables ($R^2 = 0.2831$ and $R^2 = 0.1395$, respectively), showing in this case that the distance to the champion adopter did not impact the average number of technologies adopted for each cluster.

7.3 Change from 2010 to 2013

To examine change in the network between survey years, the rand index was calculated for the 2010 and 2013 clusterings. The small rand index of 0.0061 shows that there was little change in the network from 2010 to 2013. This result is supported by the fact that across the 1122 farms and 7 technologies, only 16 technologies were adopted during this period.

8 Conclusion

Using the data from the SIMLESA survey of Ethiopia, a network which encodes the information about adoption in this country was constructed, and community structure in this network was then examined through the method of spectral clustering. As the above results show, the distance to the nearest champion adopter had no correlation to increased adoption in either individual farms or in clusters as a whole, and there was very little change in the network between 2010 and 2013, corresponding to very few technologies being adopted.

References

M. Barthélemy. Spatial networks. *Physics Reports*, 499(1-3):1–101, 2011.

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.

CIMMYT. SIMLESA overview, 2010. URL `http://simlesa.cimmyt.org`.

D. Donovan, B. Loch, H. Thompson, and J. Thompson. Change detection through clustering and spectral analysis. In *Proceedings of the International Workshop on Combinatorial Algorithms*, pages 49–58. College Publications, 2008.

S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75 – 174, 2010. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2009.11.002. URL `http://www.sciencedirect.com/science/article/pii/S0370157309002841`.

M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

C. Godsil and G. Royle. *Algebraic graph theory*. Springer, 2001.