Ethnicity and Network Centrality on the Spread of Information: A Reanalysis of Network Data From Two Ugandan Villages

Candidate Number: 1030952

Online Social Networks
Trinity 2020

Word Count: 4,990

Table of Contents

1 Introduction and Literature Review **1**

1.1 Scope of the Project 4

2 Data **6**

2.1 Descriptive Network Measures 10

3 Method and Analysis **16**

3.1 Regression Models 16

4 Results **21**

5 Conclusion and Limitations **22**

Bibliography **23**

List of Tables

2.1 Sample Count Breakdown by Variable 11
2.2 Descriptive Statistics for Network Centrality Measures 12

3.1 Logits of Different Centrality Measures (No Controls) 17
3.2 Logits for Degree Centrality 18
3.3 Logits for Eigenvector Centrality 19

List of Figures

2.1 Plot of Mugana (left) and Abalang (right) 9
2.2 Histogram for Degree Distribution 12
2.3 Histogram for Betweenness Distribution 13
2.4 Histogram for Closeness Distribution 13
2.5 Histogram for Eigenvector Distribution 14
2.6 Histogram for Page Rank Distribution 14

1 | Introduction and Literature Review

Ethnicity plays an important role in collective action as it can predict how effectively information is transmitted within a community, with implications for the political and socio-economic outcomes of a society (Chandra, 2005; Jeffrey, 2013). For example, there is evidence suggesting that greater ethnic homogeneity correlates to higher levels of economic development and public good provision in a number of developing countries (Habyarimana, Humphreys, Posner, & Weinstein, n.d.; La Ferrara, 2002; Miguel & Gugerty, 2005). In political contexts, ethnic identity shared between political candidates and voters can bring about patronage-based ethnic block voting rather than policy or programmatic-based voting (Horiwitz, 1989; Balasubramaniam, 2006).

One of the common assumptions behind this phenomenon is that ethnically homogeneous communities have denser social networks with a higher average degree compared to heterogeneous ones, thereby facilitating a better flow of information. However, there has been evidence which calls this hypothesis into question, casting into doubt the importance of network degree density in predicting the success of information transmission (Larson & Lewis, 2017). This is notably reflected in the diversity-bandwidth tradeoff argument (Aral & Van Alstyne, 2011) which postulates that a more diverse network structure increases novelty – of both persons in the network and the nature of the information – at a cost of reducing the volume of information flow. High information flow seems to be correlated with less diversity in a social network. Importantly, this can be seen to conflict Granovetter's Strength of Weak Ties argument (Granovetter, 1973) which postulates that novel information is likelier to proliferate through bridging weak social ties as opposed to stronger ones (Burt et al., 2005). In essence, while Granovetter postulates novel information flowing better through weak ties, Aral and Van Alstyne suggest it flowing better through strong ties instead.

Granovetter refers to 'bridges' between two strongly connected groups as the structural mechanism by which information is transmitted or brokered within a community. Contacts maintained through weak ties are typically unconnected to other contacts and therefore more likely have access to novel information (Granovetter, 1973). Since information in local network neighbourhoods tends to be redundant, structurally diverse contacts that reach across structural holes should provide channels through which novel information flows. For example, Burt noted that business managers who span structural holes have better evaluations, higher pay, and better ideas due to having more numerous direct and indirect (second-order and above) connections (Burt, 2004). The mechanism by which brokerage occurs is due to triadic closure. For three individuals A, B, and C: if A and B, and A and C share a strong link, triadic closure means that B and C would likely gravitate towards building a strong bond as well. Individual A would thus be said to broker the 'gap' between B and C. So, if an individual sits in between two highly connected social groups (or components), and thereby have high betweenness centrality, it is likely that they would have indirect/second-order (weak) ties to novel individuals and thus have access to their information.

Aral and Van Alstyne use a similar framework to Granovetter's conceptualisation of strong and weak ties, but with slightly different conclusions. While Granovetter and others postulate that one's position in the network is the most important predictor of the successful reception of novel information, Aral and Alstyne argue that nodal attributes and other individual characteristics are also important when considering the qualitative aspects of said information. The authors describe strong ties – with greater frequency of interaction and richer information flows – as having 'greater channel bandwidth'. Weak ties thus have a narrower bandwidth as they offer less communication, and information flows through them less frequently with lower complexity and detail (Aral & Van Alstyne, 2011). The authors use the term 'bandwidth' as it better captures the qualitative aspects of the information transmitted such as its volume.

The central argument of Aral and Van Alstyne is that there is a tradeoff between network diversity and information transmission bandwidth. For them, 'information exchange is fundamentally a social process and knowledge transfer a discretionary activity' (Aral & Van Alstyne, 2011). This means that a connection to any individual affords only the possibility of receiving the information they possess and by no means guarantees it. Other sociological contextual factors tempers the relationship between two individuals and mediates potential information transfer. So for example in competitive settings, information may even be withheld even when it is known to be in the interest of others. The authors discuss many conditions upon which information is or isn't shared; the one most applicable to this study is that of homophily – in this case most notably along ethnic lines. Unlike Granovetter, Aral and Van Alstyne bring in nodal attributes and characteristics into conceptualisations of how information is transmitted through a network. For them, it isn't just an individual's position in a network which regulates information transmission, but the characteristics of the individual also play a part in this. These characteristics can affect how, for example, trustworthy they seem to others and thereby affect the willingness of others to share information with them.

In this vein, one empirical study which explores the relationship between ethnic homophily and information transmission comes in the form of a field experiment in two villages in Uganda by Larson and Lewis (Larson & Lewis, 2017). While the original authors did not explicitly frame their work in the context of Aral and Van Alstyne's diversity-bandwidth tradeoff (Aral & Van Alstyne, 2011), their findings provide evidence for such a tradeoff to exist. In contrasting two whole-networks from two Ugandan villages, they found that despite the ethnically heterogeneous village network having a higher average degree, it was the ethnically homogeneous village network (with a lower average degree) that had the better information transmission. Ethnic homogeneity (measured by clan affiliation and linguistic markers), for Larson and Lewis, was a proxy for how trustworthy the villagers deemed each other and thereby how strong the social ties were. In this case, strong social ties seemed to be better for novel and valuable information transmission. Weak social ties, even though they were more numerous, were less useful for the transfer of novel information. They argue that the higher average degree did not matter as much for the ethnically heterogeneous village for information transmission because the villagers did not trust someone of a different ethnicity as much as they did someone of the same ethnicity, and therefore raised the probability that they withheld the information.

Even though their study emphasised ethnic homophily, Larson and Lewis also surveyed a number of other nodal attributes from the participants of the study, such as religious affiliation and sex. These may also be other reasonable dimensions for assortativity or homophily within the network (McPherson, Smith-Lovin, & Cook, 2001). Sex, or gender, is another important dimension along which strong social bonds are conditional upon. For instance, women tend to communicate more with others in their immediate familial community compared to with men (Moore, 1990), evidenced by their archetypal role as 'kinkeepers' within an extended family or community setting. The role of women as kinkeepers has been well-documented both historically (Hagestad, 1986; Rosenthal, 1985) and recently (Braithwaite, Marsh, Tschampl-Diesing, & Leach, 2017). It is also possible that religious divides can be a strong motivation for trust and distrust to form between individuals. It is not immediately clear from Larson and Lewis' study that they had rigorously controlled for these other attributes, and it is possible that homophily or assortavitity by any of these other factors would have biased their results. Any model which claims one characteristic to be the primary causal mechanism ought to control for these other aspects, and one of the aims of the study is to explicitly factor in these variables in a statistical model.

1.1 Scope of the Project

This study will be a reanalysis of the publicly available data used in Larson and Lewis' study of information proliferation in two Ugandan villages (Larson & Lewis, 2017). It aims to frame the reanalyse their findings in terms of the diversity-bandwidth tradeoff, emphasising diversity based on ethnic markers within an ego-centric network analysis framework. Because it takes an ego-centric network perspective, it will also focus on evaluating centrality measures for the combined network of the two villages. It will be guided by the research questions: (1) how strongly are different centrality measures associated with the successful transmission of information, and (2) how strongly does this correlation hold once control variables like ethnicity and gender are added?

After the introduction and literature review in Chapter 1, Chapter 2 will outline the data and method which the original Larson and Lewis study used, before explaining what their original findings were, why an ego-network framework was chosen, and what extensions this study made. Chapter 3 will analyse the strengths of the different predictors of the transmission of information by using multivariate logistic regressions, before discussing the results in Chapter 4. Chapter 5 will conclude the study.

2 | Data

The study uses publicly available data from a study by Larson and Lewis in 2017 titled 'Ethnic Networks', published in the American Journal of Political Science (Larson & Lewis, 2017). The original study investigated the association between ethnicity and the successful transfer of novel information in a rural context within a developing country, Uganda. The authors carried out their field experiment by seeding identical information in the same way at the same time in two villages that, according to them, were chosen as they were as similar as possible in all measured attributes except for ethnic composition as measured by language and clan affiliation markers.

The study described the two chosen villages as being located in the Teso region of Uganda, some 10 km north of Soroti town: Abalang in Arapai Parish, and Mugana in Aloet Parish. The two villages, Abalang and Mugana, are about two kilometres apart. The villages share a high degree of similarity in key demographic measures except that Abalang is ethnically and linguistically homogenous while Mugana is heterogenous. Both villages have about 1,400 residents, both are rural, and both are located in the same geographic region of Uganda. Mugana is a slightly wealthier, which the original study measured in terms of the quality of wall material in houses, and also has slightly more Catholic residents.

The Original Study's Data Collection and Results

The experimental data was collected in the following way by the authors first randomly seeding novel information in both villages on the same day. They seeded information with 7 randomly chosen households in Abalang and 10 in Mugana. The original study justified Mugana having more seeds because of the village's larger size. For each seed household, a Ugandan enumerator who was not from either village personally visited and shared the information by word of mouth. Mobile phone usage within the village was also not high, and where phones were used, they were predominantly for contacting persons living outside the village. This meant that the information transmission through the village was done by word of mouth. The seeded information was that in three days an event would be held at which all adults in attendance would receive a valuable block of soap in exchange for taking a survey. After the event, at least one individual in all households in both villages was surveyed to complete the measurement of social networks and the reach of information. To mitigate against enumerator differences, the two enumerators each began in a different village and then swapped villages halfway through the same day of seeding information. They then left the village and stayed away for the next three days. On the fourth day, the survey event was held at a church between and equidistant from Abalang and Mugana (Larson & Lewis, 2017).

By comparing the two whole-networks, the study found that the information spread more widely in ethnically homogenous Abalang compared to Mugana. However, this was in spite of the fact that Abalang had a less dense social network compared to Mugana. Larson and Lewis attribute this to the hypothesis that additional ties to those one trusts less were crowding out the chance to interact with more trusted people to whom information flows more freely. While they had used different terminology, this is in line with the diversity-bandwidth argument which Aral and Van Alstyne put forward. Trust, in this case, was assumed a priori to have been stronger by people of the same ethnicity and it was what the authors stated made the bigger difference in whether or not the information was transmitted.

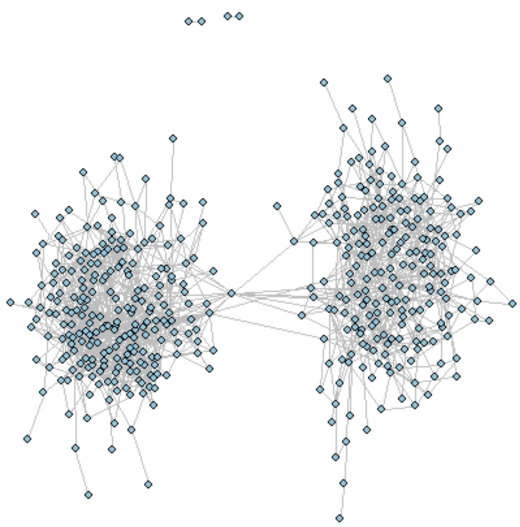
Using linguistic markers as a proxy for different Ugandan ethnicities in the villages, the original study found that the news spread more widely in ethnically homogeneous Abalang compared to ethnically heterogeneous Mugana. The news reached 62% of respondents in Abalang while only reaching 9% of respondents in Mugana. A total of 145 individuals attended the event. Only 2 of the 145 were from Mugana or elsewhere nearby while 135 were from Abalang or nearby. There were 8 individuals from villages further away. Attendees reported travelling for between 2 min to 3 hours to reach the church, with a mean travel time of 51 min. The survey at the event collected information about demographics, networks, and learning about the event. After the event, enumerators conducted surveys in both villages to collect information about demographics, networks, and knowledge of the event. Respondents who said they knew about the event but did not attend were asked what attendees received in order to verify that they in fact knew. The sampling procedure entailed first visiting all households within view of the seed households and inviting all adults to take the survey, and then visiting all other households in random order and inviting at least one adult to take the survey. A total of 226 individuals in Abalang and 237 in Mugana were surveyed in the post-event wave.

Extension and Approach of Current study

The original study's approach had contrasted the networks of Abalang and Mugana, casting them as separate networks which were independent of each other. This had allowed the original authors to draw a quasi-causal relationship as to the effect of ethnicity on information transmission. This was because a matched 'counterfactual' result was implicitly produced and compared against an 'actualised' result (i.e. information transmission in an ethnically homogeneous vs. heterogeneous network) which is one way which causality may be derived (Morgan & Winship, 2015). However, an important assumption behind this was that the networks of each of the villages were not only well-matched but also independent from each other. Even if the former is completely true, the latter claim is less solid. The original study provided two separate descriptions and plots, one each for Abalang and Mugana, which gave the impression of the independence assumption being fulfilled. In pre-processing the data, however, this study took the two graphs representing each network and combined them into one network graph. This results in the following plot shown in figure 2.1 which reveals that the two villages are in fact not independent from each other. There were at least 8 edges flowing from one village to the other which meant that information could have easily flowed from one village to the other. Information transmission in one village is therefore not independent of the other; and there are therefore issues of edogeneity present biasing the original results.

Figure 2.1 reveals the presence of social actors who bridge the gap between the two villages and visually shows how the two networks are not independent. While this does not necessarily invalidate the findings of the previous study, it calls into question the validity of the original study's results.

Figure 2.1: Plot of Mugana (left) and Abalang (right)



To avoid the pitfalls of the original study, this study will take an ego-network approach to analyse the same research question of whether ethnicity is associated with the likelihood of successful information transmission. Importantly, it also considers whether a node's position in the network is also a significant factor. This study will re-analyse the data that the original study collected, investigating how strongly two sets of factors are correlated with successful information transmission: (1) one's ego-network characteristics, and (2) one's nodal characteristics such as ethnicity, gender, age, or religion etc.

This study seeks to compare ego-networks by looking at individual-level network centrality metrics. After calculating and describing the network characteristics, it then will perform a multivariate logistic regression with the main outcome variable being whether the person has heard of the event and the main predictor variable being the various centrality measures. This study will then add in other secondary predictor variables such as ethnicity, gender, and other appropriate characteristics available in the data. This study does not explicitly analyse whether or not individuals have attended the event.

2.1 Descriptive Network Measures

The original study gave a description and breakdown of the variables at the village level, analysing and describing Abalang and Mugana separately. Since this study will analyse the two villages as an amalgamated network, it will not repeat the sample counts by village. Instead, it will provide descriptive statistics on an aggregated level with all 512 respondents who were living in both Abalang or Mugana. The original study used post-event surveys and used closed networks so only links between survey respondents were included. Persons nominated by a survey respondent who were themselves not also survey respondents were excluded. This is shown in table 2.1. Frequencies may not always total to 512 due to some node attributes containing missing values. The following figures will show different measures of centrality and how they are distributed in the graph. Table 2.2 will then show measures of central tendency for each of the 5 centrality measures of degree, betweenness, closeness, eigenvector, and page rank. Figures 2.2 to 2.6 will then display this information visually using histograms.

Figure 2.2 will show the degree distribution in the network. Degree centrality is an importance score based simply on the number of links held by each node. It tells us how many immediate connections each node has to other nodes in the network. Most nodes have a degree of 10 or fewer while few nodes have a higher degree, forming a distribution that has a right skew. Figure 2.3 shows the distribution of betweenness scores. Betweenness centrality measures the number of times a node lies on the shortest path between other nodes. It shows which nodes bridge structural holes between nodes in a network and does this by identifying all the shortest paths and then counting how many times each node falls on one. It shows a power law distribution which is extremely right skewed.

Table 2.1: Sample Count Breakdown by Variable

Variable	Frequency	Percentage
Heard about event		
Yes	203	40.5%
No	298	59.5%
Attended event		
Yes	69	13.8%
No	431	86.2%
Respondent sex		
Male	165	32.6%
Female	341	67.4%
Respondent's village of residence		
Abalang	271	52.9%
Mugana	241	47.1%
Respondent's clan		
Ikaribwok	164	37.1%
Ikatekok	69	15.6%
Imugana	14	3.2%
Irarak	71	16.1%
Others	124	28.0%
Is respondent married		
Yes	427	85.9%
No	70	14.1%
Language spoken with family		
Ateso	410	82.7%
Kumam	79	15.9%
Other	7	1.4%
Employment status and type		
No paid work, not looking for job	198	39.8%
No paid work, looking for job	19	3.8%
Has paid work, part-time	219	44.1%
Has paid work, full-time	61	12.3%
Highest educational qualification		
No formal schooling	83	16.7%
Informal schooling only	8	1.6%
Some primary schooling	239	48.1%
Completed primary school	50	10.1%
Some secondary/high schooling	80	16.1%
Completed secondary/high school	21	4.2%
Tertiary non-degree qualification	16	3.2%

Variable	Frequency	Percentage
Respondent's religion		
Catholic	372	75.3%
Anglican	61	12.3%
Muslim	2	0.4%
Pentecostal	56	11.3%
Other	3	0.6%

Variable (Continuous)	Mean	Std Dev.
Age	38	16.81

Table 2.2: Descriptive Statistics for Network Centrality Measures

Centrality Measure	Median	Mean	Standard Deviation
Degree	6.00	7.36	5.08
Betweenness	574.45	2038.02	5016.76
Closeness	0.00026	0.00026	0.00003
Eigenvector	0.002	0.063	0.142
Page Rank	0.00136	0.00228	0.00278

Figure 2.2: Histogram for Degree Distribution

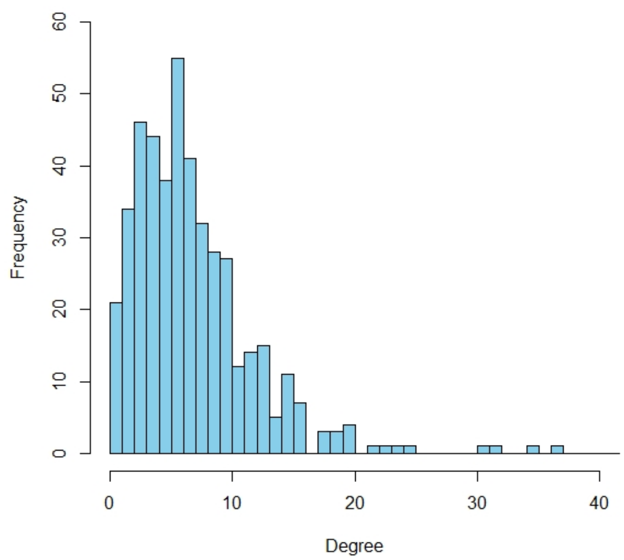


Figure 2.3: Histogram for Betweenness Distribution

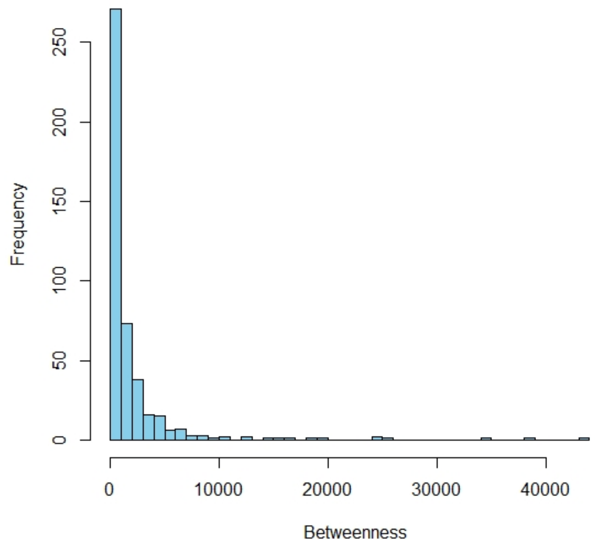


Figure 2.4: Histogram for Closeness Distribution

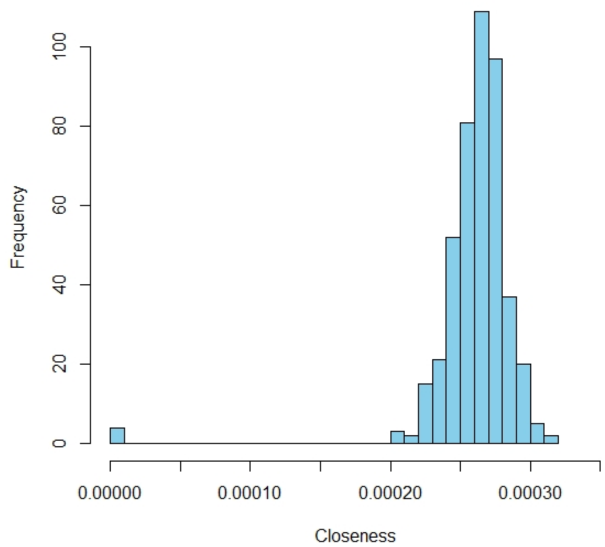


Figure 2.5: Histogram for Eigenvector Distribution

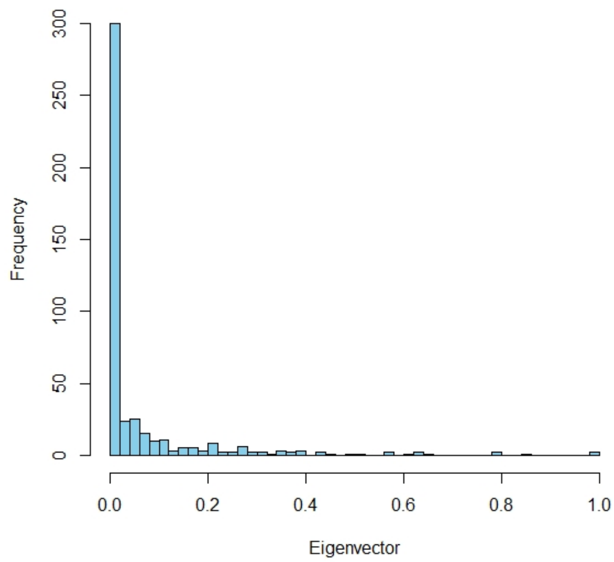


Figure 2.6: Histogram for Page Rank Distribution

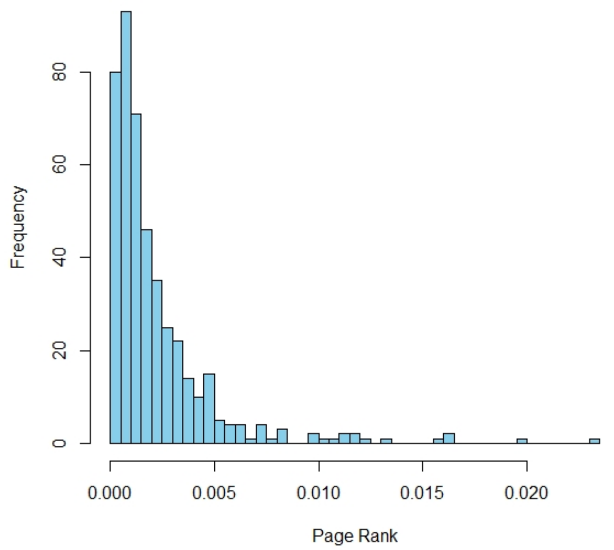


Figure 2.4's closeness scores calculates the shortest paths between all nodes, then assigns each node a score based on its sum of shortest paths. It is the most symmetrically distributed of all the centrality measures except for the 4 cases of nodes which are disconnected from the rest of the graph, earning a closeness score of 0. Figure 2.5 shows the distribution of eigenvector centrality. It measures a node's influence based on the number of links it has to other nodes. It then taking into account how well connected a node is, and how many links their connections have, and so on. It is distributed in a power law fashion like for betweenness centrality.

Lastly, figure 2.6 shows page rank scores. Page rank is a variant of eigenvector centrality, which measures a node's influence based on the number of links it has to other nodes in the network. It also assigns nodes a score based on their connections, and their connections' connections. This measure gives us information on nodes whose influence extends beyond their direct connections into the wider network. Like betweenness centrality, most nodes have a low page rank score while a few nodes are very highly connected thus forming a pareto or power law distribution (though the skew is not as strong as for betweenness).

All these measures will be used as different independent variables to ascertain which is the best predictor of successful information transmission in the given context. Due to issues of multicollinearity, the different measures of centrality will not be used together in the same regression model as covariates.

Finally, further justifications of the inappropriateness of separating the networks of Abalang and Mugana show themselves when describing the path lengths of the networks. The networks of Abalang and Mugana both have a diameter of 13 each, but when combined, the diameter is 17. The average path length of Abalang is 5.34 and Mugana's is 4.53, while the combined average path length is 6.76. The relatively minor increases of the aggregated path metrics compared to the corresponding individual metrics suggests that these two networks cannot be treated independently of each other since there are a considerable overlap of social ties between the villages. Despite both villages having a diameter of 13, the combined total diameter is only increased by 4 to a value of 17.

3 | Method and Analysis

The data wrangling and analysis for this paper was done with open-sourced programming language R (version 3.6.2), primarily drawing on the igraph package (Csardi & Nepusz, 2006). The original study also used R and thus it was possible to reuse the some of the code the authors had provided where appropriate such as in the reconstruction of the network graphs.

The key outcome variable to be analysed is whether a respondent had heard of the event. The key predictor variables which will be used separately are the different centrality measures: degree, betweenness, closeness, eigenvector, and page rank. They will not be included together in the same regression model because of the high multicollinearity that will exist between them since they are derived in a similar manner and reflect similar network properties. It would be inappropriate to use them simultaneously as 5 variables in one regression model. The covariates used (reflected in table 2.1) are as follows: age, sex, village of residence, clan affiliation, language spoken, marital status, employment, educational qualification, and religion.

3.1 Regression Models

A logistic regression model is chosen because the outcome variable is a binary one: whether a respondent has or has not heard of the event, represented below by variable $Heard$. The logistic regression models used will take the following form:

$$log(\frac{P(Heard)}{1-P(Heard)}) = \beta_0 + \beta_1 Centrality + \beta_2 Cov_a + \beta_3 Cov_b + ... + \epsilon$$

Where $P(Heard)$ is the probability that a respondent has heard of the event, and thus $\frac{P(Heard)}{1-P(Heard)}$ is the odds of hearing of the event. The β_n terms are therefore logits – taking exponent of a logit will result in an odds ratio being produced. $Centrality$ represents the centrality measure used. The Cov_n terms represent the combinations of covariates/control variables which may be included in iterations of different models. β terms are coefficients of the variables and ϵ is the error term.

One of the weaknesses of this approach is the relatively low statistical power of the models due to low sample counts. Therefore, this study did not begin by omitting all cases with any missing variables. While this maximises the statistical power in models with fewer variables (and thereby a lower rate of missingness), there is the tradeoff of having an unequal N (and thus statistical power) across statistical models which may slightly bias some of the results.

Table 3.1 shows the value of the logits when regressing $Heard$ on different centrality measures, with no controls used.

Table 3.1: Logits of Different Centrality Measures (No Controls)

Centrality Measure	Logit	SE	p-value
Degree	-0.018	0.020	0.360
Closeness	-0.002	0.003	0.488
Betweenness	0.00002	0.00002	0.292
Eigenvector	-5.610	1.445	<0.001
Page Rank	27.395	34.258	0.424

The table shows that centrality measures, unaided by any other variables, do a poor job of evidencing how network positions increase the odds that a respondent would hear of the event due to their high p-values. Eigenvector centrality is the sole exception in this case, although its logit of -5.61 corresponds to an odds ratio of 0.004 suggesting a negative association.

The next part of the study tests the strength of association between the different centrality measures once different controls are added. These controls will be grouped into three categories: (1) ethnic markers (clan affiliation and language spoken), (2) socio-demographic variables (gender, marital status, employment, educational qualification, religious affiliation, and age), and (3) village of residence. There will be 7 models run for each centrality measure. Only degree and eigenvector centrality yield statistically significant results at the 10% level. These are reflected in tables 3.2 and 3.3.

Table 3.2: Logits for Degree Centrality

Models:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β_0 (Constant)	-0.450***	-0.232	-2.287***	0.262	0.317	-1.962***	-0.964
Degree	-0.018	0.001	-0.024	0.047*	0.053**	-0.014	0.040
Ikaribwok clan							
Ikatekok clan		0.320			-0.008	0.288	0.085
Imugana clan		-2.259**			-0.065	-1.792*	0.378
Irarak clan		-0.883**			-0.902**	-0.975**	-0.822*
Other clan		-0.068			0.078	-0.065	0.059
Ateso language							
Kumam language		-1.626***			0.324	-1.482***	0.340
Other language		-0.438			-1.022	-0.098	-0.023
Female							
Male			-0.601**			-0.406	-0.512
Unmarried							
Married			0.826**			0.680*	0.991**
No job, not looking							
No job, looking			-0.567			-0.410	-0.330
Part-time job			0.156			0.197	-0.073
Full-time job			-0.302			-0.631	-0.805
No schooling							
Informal schooling			1.059			0.869	0.474
Some primary schooling			0.597*			0.653*	0.733
Completed primary school			0.038			-0.018	0.273*
Some high school			1.098**			1.353***	1.564***
Completed high school			1.101*			1.492**	1.198
Tertiary non-degree			0.436			0.567	0.195
Catholic							
Anglican			0.403			0.339	-0.176
Muslim			-13.446				
Pentecostal			1.473***			1.049***	0.014
Other religion			1.279			0.896	0.899
Age			0.013*			0.0129	0.003
Abalang village							
Mugana village				-2.945***	-3.087***		-3.220***
N	455	400	438	455	400	391	391

*** < 0.01 ** < 0.05 * < 0.1

Table 3.2 shows that without controls (model 1), there is insufficient evidence for an association between degree centrality and hearing of the event. Only by adding in ethnicity and village of residency (model 5) does the level of significance reach an appropriate level for a significant correlation to be derived. Controlling for village of residence (model 4) is also sufficient, albeit not as strong, but controlling for ethnicity alone (model 2) is insufficient. In all models, marital status and clan affiliation was significant. Respondents' religious affiliation was also significant but not when village of residence was also controlled for.

Table 3.3: Logits for Eigenvector Centrality

Models:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β_0 (Constant)	-0.334***	-0.072	-2.361***	0.550***	0.670***	-2.095***	-0.805
Eigenvector	-5.610***	-4.353***	-4.551***	0.794*	0.646	-3.860***	0.688
Ikaribwok clan							
Ikatekok clan		0.268			-0.046	0.309	0.061
Imugana clan		-2.218**			-0.164	-1.797*	0.318
Irarak clan		-0.731**			-0.876**	-0.845**	-0.834*
Other clan		-0.098			0.004	-0.058	0.012
Ateso language							
Kumam language		-1.374***			0.360	-1.261***	0.362
Other language		-0.625			-1.125	-0.149	-0.062
Female							
Male			-0.467*			-0.334	-0.524
Unmarried							
Married			0.855**			0.744*	1.038**
No job, not looking							
No job, looking			-0.593			-0.446	-0.303
Part-time job			0.147			0.187	-0.068
Full-time job			-0.294			-0.575	-0.842*
No schooling							
Informal schooling			0.763			0.595	0.407
Some primary schooling			0.570*			0.621*	0.732*
Completed primary school			0.067			0.010	0.273
Some high school			1.129**			1.417***	1.652***
Completed high school			1.001*			1.3256*	1.260
Tertiary non-degree			0.291			0.460	0.216
Catholic							
Anglican			0.335			0.309	-0.165
Muslim			-13.114				
Pentecostal			1.299***			0.937**	0.029
Other religion			1.121			0.815	0.809
Age			0.016**			0.016*	0.004
Abalang village							
Mugana village				-2.914***	-3.023***		-3.192***
N	455	400	438	455	400	391	391

*** < 0.01 ** < 0.05 * < 0.1

Table 3.3 shows eigenvector centrality being a significant predictor of hearing of the event (at the 5% level) only when village of residence is not controlled for (models 1, 2, 3, and 6). This association is also negative (negative logits translating to an odds ratio of less than 1, or probabilities of less than 0.5) which actually means that, without controlling for one's locale, nodes that have more eigenvector centrality are *less* likely to receive the information. The direction of association is reversed once village is controlled for (model 4), showing a positive and significant relationship (albeit at the 10% level) instead.

Comparing the two centrality measures, tables 3.2 and 3.3 show differing circumstances that degree and eigenvector centrality become statistically significant predictors of successful information transmission. For instance, degree centrality is not significant when there are fewer controls (models 1 to 3) used whereas eigenvector centrality is highly significant under those circumstances. Eigenvector centrality remains significant when many variables are controlled for, such as in model 6 controlling for ethnicity and socio-demographic variables. On the contrary, degree centrality is not significant once these variables are controlled for.

Degree centrality is only statistically significant at the 10% level once village residency is factored in (Model 4) and significant at the 5% level when ethnic markers are also included (Model 5). Degree centrality seems to work in tandem with ethnic markers and village residency, and the relationship between degree centrality and the odds of information transmission is positive. The more edges a node has, controlling for ethnicity and locale, the higher the odds of the individual successfully receiving the information.

In contrast, eigenvector centrality remains highly significant across all statistical models except when both village residency and ethnicity is factored in. However, the logits taking a negative value means a negative association between eigenvector centrality and the odds of hearing of the event. This means that the more well-connected a node is to others in the network, the less likely they are to hear of the event. A further discuss of this apparent paradox will be included in the next chapter.

A final descriptive observation is that when all control variables representing ethnic markers, socio-demographic variables, and village residency were incorporated into the regression model, logits for both degree and eigenvector centrality become statistically insignificant. The only variables which this model consistently showed as being significant was village residency, marital status, clan affiliation (notably, the Irarak clan when compared to the Ikaribwok clan), and having some high school education (when compared against respondents who had no education). Socio-demographic variables seem to be more powerful at predicting whether a respondent has heard of the event.

4 | Results

The results of this analysis paint a counter-intuitive picture: while we expect degree centrality to positive correlate with the odds of hearing of the event, eigenvector centrality ends up correlating negatively instead. This is the case with or without controls. However, they can be explained in the context of the diversity-bandwidth argument (Aral & Van Alstyne, 2011).

At its core, the argument postulates that a greater number of ties between less trusting individuals impedes the successful transmission of information. In this study, as did the original, trust was assumed to have been present between individuals who shared clan affiliation and language. The logit for degree centrality (although statistically insignificant) took a negative value when no controls were used, but became positive (and significant) once ethnic markers and village locale were explicitly taken into account. Once included in the model, the logits for clan affiliation and language in models 2 and 5 (table 3.2) were highly negative, showing that differences in ethnic markers diminished the odds of hearing of the event. The diverging directions of the logits of degree centrality and ethnic markers/village residency demonstrates that information did not tend to flow between persons who did not share common characteristics (and were assumed to trust each other less) as predicted by the diversity-bandwidth tradeoff argument.

While the negative logits for ethnic markers were also evident in the regressions for eigenvector centrality, what is harder to account for is why the logit for eigenvector centrality remains negative even after ethnic markers, village locale, and socio-demographic variables are controlled for. It is feasible this has to do with the low statistical power of this particular regression, or that the overall negative effect of ethnic heterogeneity outweighs the positive of ethnic homogeneity in this network. Regardless, the logits become increasingly positive as ethnic markers and village of residency are incrementally controlled for which is still in line with the above diversity-bandwidth tradeoff framework.

5 | Conclusion and Limitations

The diversity-bandwidth tradeoff is a powerful explanatory framework which is signficant as it adds an additional layer of analysis which goes beyond network centrality measures in accounting for the successful transmission of novel information. This study successfully evidences this framework and also does not refute the original Larson and Lewis study in their findings. This study, like theirs, also find that ethnic homogeneity raises the odds that information is likely to be proliferated albeit done from a different network perspective i.e. this study's ego-network versus Larson and Lewis' whole-network analysis.

One limitation of this study is that while datasets containing a large number of networks permit comparisons of group-level outcomes with high statistical power, comparing the network features in this data may be problematic due to the small sample of nodes. Because a network's nodes are systematically related to one another, small random samples of nodes do not necessarily reveal properties of a larger network in the same way that random samples of independent observations reveal features of the broader population from which they are drawn from (Lee, Kim, & Jeong, 2006). This result may therefore not be generalisable to larger networks in other contexts. A possible extension would therefore be to replicate this study on a larger dataset to achieve better results.

Bibliography

Aral, S., & Van Alstyne, M. (2011). The diversity-bandwidth trade-off. *American journal of sociology*, *117*(1), 90–171.

Balasubramaniam, V. (2006). Embedding ethnic politics in malaysia: Economic growth, its ramifications and political popularity. *Asian Journal of Political Science*, *14*(1), 23–39.

Braithwaite, D. O., Marsh, J. S., Tschampl-Diesing, C. L., & Leach, M. S. (2017). 'love needs to be exchanged': A diary study of interaction and enactment of the family kinkeeper role. *Western Journal of Communication*, *81*(5), 601–618.

Burt, R. S. (2004). Structural holes and good ideas. *American journal of sociology*, *110*(2), 349–399.

Burt, R. S., et al. (2005). *Brokerage and closure: An introduction to social capital*. Oxford university press.

Chandra, K. (2005). Ethnic parties and democratic stability. *Perspectives on politics*, *3*(2), 235–252.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695. Retrieved from `http://igraph.sf.net`

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*.

Habyarimana, J., Humphreys, M., Posner, D. N., & Weinstein, J. M. (n.d.). Coethnicity: Diversity and the dilemmas of collective action. *Social Forces*, *89*, 4.

Hagestad, G. O. (1986). The family: Women and grandparents as kinkeepers. *Our aging society: Paradox and promise*, 141–160.

Horiwitz, D. L. (1989). Incentives and behaviour in the ethnic politics of sri lanka and malaysia. *Third World Quarterly*, *11*(4), 18–35.

Jeffrey, C.-K. (2013). Information and ethnic politics in africa. *British Journal of Political Science*, *43*(2), 345–373.

La Ferrara, E. (2002). Self-help groups and income generation in the informal settlements of nairobi. *Journal of African Economies*, *11*(1), 61–89.

Larson, J. M., & Lewis, J. I. (2017). Ethnic networks. *American Journal of Political Science*, *61*(2), 350–364.

Lee, S. H., Kim, P.-J., & Jeong, H. (2006). Statistical properties of sampled networks. *Physical review E*, *73*(1), 016102.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415–444.

Miguel, E., & Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in kenya. *Journal of public Economics*, *89*(11-12), 2325–2368.

Moore, G. (1990). Structural determinants of men's and women's personal networks. *American sociological review*, 726–735.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press.

Rosenthal, C. J. (1985). Kinkeeping in the familial division of labor. *Journal of Marriage and the Family*, 965–974.