Assessing the Impact of Socio-Economic HIV Driving Factors in Ghana: Rural Versus Urban*

[1]Agangiba W.A and [1]Agangiba M.
[1]University of Mines and Technology (UMaT), Tarkwa Ghana

Agangiba, W. A. and Agangiba, M. (2019), "Assessing the Impact of Socio-economic HIV Driving Factors in Ghana: Rural Versus Urban", *Ghana Journal of Technology,* Vol. 3, No. 2, pp. 42 - 50.

Abstract

For several decades, HIV/AIDS has been an important concern to key international and intergovernmental organisations and a subject of many academic research efforts. Since its emergence in the early 1980s, the devastating effects of the HIV/AIDS pandemic on human life across the world cannot be overemphasised. In the academic front, many researchers have identified important socio-economic drivers of the pandemic, particularly, in sub-Saharan Africa which is the most hard-hit. Some of the socio-economic drivers identified include educational level, marital status and occupation. However, beyond the identification of these factors, not much effort has been made to assess their degree of impact on the lives of a given population. This paper contributes to fill this gap in literature by assessing the degree of impact of such factors. Assessing the degree of impact of such factors in numerical measures is imperative not only to compare the extent of the impact of individual factors but more importantly to facilitate the development of policies to counter the devastating effects of the pandemic. Using data mining approach, this paper assesses the impact of HIV/AIDS driving factors and compares the impact of such factors in the urban areas with their impact in the rural areas in Ghana. The results show that some driving factors which are very important in the rural settings do not constitute significant driving factors in the urban settings and vice versa.

Keywords: HIV/AIDS, Pandemic, Socio-Economic, Degree of Impact, Driving Factors

1 Introduction

Acquired Immunodeficiency Syndrome (AIDS) is caused by Human Immunodeficiency Virus (HIV) and was first detected in 1981 (Sharp and Hahn, 2011). In 1984, a study revealed that a virus known as HIV was the cause of AIDS (Blattner et *al*., 1988). The virus comes in two varieties, namely HIV-1 and HIV-2 with the former being the commonest type transmitted through the exchange of fluids (Arrehag *et al.*, 2006; Heeney, Dalgleish and Weiss 2006, Shar and Hahn 2010). Even though sexual activity is known as the primary means by which the virus is transmitted from person to person, it could also be spread through percutaneous and perinatal procedures (Hladik and McElrath 2008; Cohen *et al.*, 2011).

World statistics have consistently highlighted Africa (particularly sub-Saharan Africa) as the most-burdened with HIV/AIDS with approximately 70% of world's HIV/AIDS cases recorded in the region (Were and Nafula, 2003; Arrehag et *al.*, 2006; Kaya, 2018). Further, studies show that 1 out of every 20 persons in the region is infected with HIV (Biney *et al.*, 2015). Given that sexual behaviour is a major mode of HIV/AIDS infection, more economically active age groups are usually at higher risk of contracting the disease since they are sexually very active. As a result, economic growth of countries such as those in sub-Saharan Africa, where HIV/AIDS is most prevalent, is greatly affected. Morbidity and inability of HIV/AIDS patients to work pose a lot of economic hardships to families and financial losses to business and corporate organisations, thereby leading to increase in poverty and orphanages across Sub-Saharan Africa (Arrehag *et al.*, 2006). Subsequently, HIV/AIDS has been termed a socio-economic pandemic (Kaya, 2018). For this reason, tackling the pandemic within any context would require both social and economic efforts alongside strong political will (Patterson, 2018).

Several research works have been conducted across the African continent, regarding the identification of socio-economic factors that influence the spread of HIV/AIDS on the continent. For instance, Nagoli *et al. (*2010), using logistic regression identified low education level and occupation as drivers of the pandemic in Malawi. Similarly, Bogale *et al*. (2009), identified low education as a driving factor of the pandemic in Ethiopia in their study using frequency distribution, cross-tabulation and chi-square analyses. Using retrospective ecological comparison, trend analysis and chi-square test, Parkhurst (2010), showed that both poverty and wealth are drivers of the pandemic across selected African countries. The study highlighted premarital sex, multiple sexual partners and poverty as key drivers of the pandemic. In another study conducted by Fortson (2008), across Burkina Faso, Cameroon, Ghana, Kenya and Tanzania, the findings showed that wealth was associated with HIV/AIDS positive status. Other factors associated with HIV/AIDS infection across sub-Saharan Africa include unemployment,

**Manuscript received November 24, 2018*
Revised version accepted March 20, 2019

disrupted marriage (widowhood and divorce), and urban lifestyle including access to media such as radio, television and newspapers (Isiugo-Abanihe, 2004; Kalichman, 2006; De Walque, 2009). These studies though identified socio-economic drivers of HIV/AIDS, they did not shed light on the level of impact of each driver. As a result, it becomes difficult to assess socio-economic drivers of HIV/AIDS in order of importance to inform intervention or measures needed from decision makers. Detail assessment of the driving factors would guide policymakers and planners in their efforts to curb the pandemic more effectively.

This paper focuses on Ghana, a country in sub-Saharan Africa. Studies show that HIV/AIDS prevalence in Ghana reduced from 3.6 per cent in 2003 to 1.4 per cent as at 2015 (Biney *et al.,* 2015; Ankrah *et al.,* 2017). The reduction in the rate of prevalence could be attributed to strategies implemented by the government over the years to control the disease including sponsoring treatment of infected persons (Ankrah *et al.,* 2017). Consequently, knowing the impact of HIV/AIDS socio-economic drivers would further assist government to formulate right policies and improve strategies to control the disease.

This paper uses clustering; a data mining technique to assess the degree of impact of HIV/AIDS socio-economic drivers in Ghana. The study further compares the results obtained from the urban areas to that of the rural areas.

2 Resources and Methods Used

2.1 Data Collection

The dataset used for this research is an anonymous secondary data obtained from the National AIDS/STI Control Programme (NACP). NACP is the body mandated for HIV/STI sentinel surveys in Ghana. The attributes constituting the dataset are Age, Gender, Marital Status, Education Level and Occupation. For each record, the attributes are characterised by their respective categories or nominal values. For instance, Marital Statuses (MS) is characterised by *Married (M), Widow(er) (W), Cohabiting(C), Single* (S) and *Separated(Sep)*. Values of the Education Level (E) include Primary (P), Tertiary(T), Nil (N) and Junior Secondary School (JSS).

The original data contained more than 200 individual occupations such as blacksmith, teacher, farmer, etc. For this paper, however, such occupations were classified into 9 categories using the International Standard Classification of Occupations 2008 (Anon, 2012). Table 1 shows a sample of the dataset.

Table 1. Sample data

MaritalStats	EduLevel
Widow	Primary
Widow	Primary
Single	Primary
Single	Tertiary
Married	Tertiary
Widow	JSS
Married	JSS
Married	JSS
Married	JSS
Married	JSS

2.2 Processing of data: Binarization

This first step in the modelling process converts all categorical data units into numerical data through the process of binarization. Given a dataset D with n attributes and m records (rows), let each attribute be composed of a finite number of possible nominal values. If the set of all distinct nominals of all attributes in D is $k = \{k_1, k_2, . . . , k_n\}$ then each record can be converted to a binary string of length n by the following scheme:

$$f(k_i)=\begin{cases} 1 & ;\text{if } k_i \text{ is present} \\ 0; & \text{otherwise} \end{cases} \quad (1)$$

For instance, the binarized form of Table 1 is shown in Table 2. Each row is a binary string (vector) of 0 and 1.

Table 2 Binarized form of Sample Data

MS=W	MS=S	MS=M	E=P	E=T	E=JSS
1	0	0	1	0	0
1	0	0	1	0	0
0	1	0	1	0	0
0	1	0	0	1	0
0	0	1	0	1	0
1	0	0	0	0	1
0	0	1	0	0	1
0	0	1	0	0	1
0	0	1	0	0	1
0	0	1	0	0	1

2.3 Clustering the Data

The second major step in the process is clustering; an unsupervised data mining technique. Clustering is a mathematical technique of dividing a dataset into groups of similar objects (Berkhin, 2006; Sehgal and Garg, 2014). The common classes of clustering algorithms are partition, hierarchical or density Based (Saraswathi and Sheela, 2014; Sisodia et al., 2012). In this paper, however, *Growing Neural Gas (GnG)* (Fritzke, 1995) is used. GnG is a topology preserving clustering technique, belonging to a class of techniques called unsupervised competitive learning neural networks (Pena et al., 2008; Ngo *et al.*, 2014). Details of GnG is described in (Fritzke, 1995). For instance, GnG splits Table 2 into three clusters: C0, C1and C3 as shown in Table 3.

Table 3 Clusters form from sample data

MS=W	MS=S	MS=M	E=P	E=T	E=JSS	
1	0	0	1	0	0	
1	0	0	1	0	0	**C0**
0	1	0	1	0	0	
0	1	0	0	1	0	
0	0	1	0	1	0	**C1**
1	0	0	0	0	1	
0	0	1	0	0	1	
0	0	1	0	0	1	
0	0	1	0	0	1	**C2**
0	0	1	0	0	1	

2.4 Assessment of Degree of Impact

The paper adapts the concept of Feature Maximisation for the assessment of degrees of interactivities of HIV drivers. Feature Maximisation is proposed by Lamirel and Al Shehabi, (2015) as a cluster quality assessment method. It is highly suitable for clustering high dimensional datasets. Furthermore, Feature Maximisation is non-parameterised and less sensitive to noise; thereby making it independent of the clustering method being used. To determine the degree of impact of a given driving factor, key indices are defined as follows for each driving factor per cluster as follows:

Feature Recall (FR): This is the ratio of the weight (sum of non-zero appearances) of a factor in a given cluster to its weight across all clusters. Taking the set *C* of all clusters formed and for each data (binary string) *d*, in a given cluster *c*, mathematically **FR** is expressed as shown in equation 1.

$$FR_c(f)=\frac{\sum_{d\in c} W_d^f}{\sum_{c\in C}\sum_{d\in c} W_d^f} \qquad (1)$$

W_d^f is the weight of the factor *f*. For instance, the weight of *MaritalStats=Widow* in cluster C0 is 2 and its weight across the set of clusters formed is 3. Therefore $FR_c(Widow)=\frac{2}{3}=0.66$.

Feature Predominance (FP): This is the ratio of the weight (sum of non-zero appearances) of a factor in a given cluster to the sum of weights of all other factors in the same cluster. **FP** is expressed mathematically as shown in equation 2.

$$FP_c(f)=\frac{\sum_{d\in c} W_d^f}{\sum_{f'\in F_c,\ d\in c} W_d^{f'}} \qquad (2)$$

The sum of weights of all factors in cluster C1, for example, is 6. $FP_c(Widow)=\frac{2}{6}=0.33$.

Feature F-Measure (*FFc*): *FFc* of a feature *f* is the harmonic mean of the feature recall and the feature predominance.

$$FF_c(f)=2\left(\frac{FR_c(f)*FP_c(f)}{FR_c(f)+FP_c(f)}\right) \qquad (3)$$

Therefore, $FF_c(Widow)=2\left(\frac{0.66*0.33}{0.66+0.33}\right)=0.44$

Average F-Measure of *f*: Average F-measure of *f* is the sum of all F-measures of *f* in the clusters c' where *f* appears divided by the number of clusters in which *f* appears. This is computed as follows:

$$\overline{FF}(f)=\sum_{c'\in C}\frac{FF_{c'(f)}}{|C_{/f}|} \qquad (4)$$

$C_{/f}$ is the subset of *C* where *f* occurs.

Average F-Measure of all Factors in *C*: Average F-measure of all the features across the set of clusters formed is the sum of F-measures of all feature divided by the total number of features across all clusters. This is computed as follows:

$$\overline{FF}_D=\sum_{f\in F}\frac{\overline{FF}(f)}{|F|} \qquad (5)$$

F is the total number of features in the *C*. A factor *f* is considered significant and retained in the given cluster as an important driver of the pandemic if it meets the following condition:

$$\{FF_c(f)>\overline{FF}(f) \text{ and } FF_c(f)>\overline{FF}_D\} \qquad (6)$$

Any factor not respecting the second condition (equation 6) in any cluster are ignored.

Contrast: Contrast is defined as the ratio between the *F*-measure of *f* (i.e. *FF*c(*f*)) and the average *F*-measure *FF* of *f* for the whole set of formed clusters as expressed in equation 7.

$$G_c(f)=\frac{FF_c(f)}{\overline{FF}(f)} \qquad (7)$$

Contrast is the indicator of the performance or impact of each retained factor *f*. Active features are those retained factors for which contrast is greater than 1.

3 Results and Discussion

The datasets were executed in turns through the processes of binarization, clustering and feature maximization. The clusters being the ultimate outcome are presented and discussed here. Sixteen clusters were formed from Rural dataset (rural setting) while ten clusters were obtained from the Urban dataset (urban setting). Cluster 8 from the Rural dataset is empty. The numerical value preceding the element in each cluster is the contrast of that element. For instance, in *Cluster 0* of the Rural HIV dataset, the contrast of Age Group=A30_34 is 5.903213.

The following are clusters formed from the rural dataset:

Cluster 0

5.903213 Age Group=A30_34
2.307325 Occupation =Service and Sales Workers
2.185948 Marital Status=Married
2.105852 Education Level=Nil
2.001706 Gender=F
1.972840 Education Level=Primary
1.941243 Education Level=Middle
1.732292 Occupation =Agricultural, Forestry and Fishery Workers
1.406363 Occupation =Other Occupations
1.307975 Marital Status=Single

Cluster 1

8.307237 Occupation =Clerical Support Workers
7.246994 Occupation =Elementary Occupations

Cluster 2

4.472543 Age Group=A15_19
4.081807 Age Group=A60_64
3.376286 Age Group=old_Age
2.919618 Age Group=A0_15
2.086554 Marital Status=Single
1.698468 Occupation =Other Occupations

Cluster 3

3.608029 Marital Status=Separated

Cluster 4

3.778031 Marital Status=Cohabiting

Cluster 5

5.155208 Age Group=A25_29
2.003674 Occupation =Service and Sales Workers
1.942867 Gender=F
1.877045 Education Level=Primary
1.808372 Marital Status=Married
1.787706 Occupation =Other Occupations
1.742342 Marital Status=Single
1.715227 Education Level=Middle
1.664597 Education Level=Nil
1.153323 Occupation =Agricultural, Forestry and Fishery Workers

Cluster 6

5.131315 Age Group=A35_39
2.145078 Education Level=Nil
1.972653 Marital Status=Married
1.938738 Occupation =Agricultural, Forestry and Fishery Workers
1.816797 Occupation =Service and Sales Workers
1.806447 Education Level=Primary
1.580132 Education Level=Middle
1.272851 Occupation =Other Occupations

Cluster 7

4.570749 Marital Status=Widow(er)
1.807060 Age Group=old_Age
1.694989 Age Group=A60_64
1.506089 Education Level=Nil
1.052185 Age Group=A40_44

Cluster 9

7.309581 Occupation =Crafts and Related Trade Workers
1.622901 Education Level=Middle
1.329916 Marital Status=Single
1.080797 Age Group=A25_29
1.051032 Marital Status=Married

Cluster 10

9.635889Education Level=Post Secondary/Tertiary
4.669693 Occupation =Professionals
2.232677 Education Level=Secondary

Cluster 11

5.146141 Age Group=A40_44
1.748569 Education Level=Nil
1.673849 Marital Status=Divorced
1.664439 Occupation =Agricultural, Forestry and Fishery Workers
1.500695 Education Level=Primary
1.493813 Marital Status=Married
1.276561 Occupation =Service and Sales Workers
1.193188 Education Level=Middle

Cluster 12

2.861619 Age Group=A55_59
1.350237 Marital Status=Divorced

Cluster 13

4.343986 Age Group=A20_24
1.935770 Marital Status=Single
1.663598 Occupation =Other Occupations
1.496559 Education Level=Primary
1.207336 Education Level=Middle
1.133837 Marital Status=Married
1.074076 Occupation =Service and Sales Workers
1.010120 Occupation =Crafts and Related Trade Workers

Cluster 14

4.809729 Age Group=A45_49
1.857202 Marital Status=Divorced
1.488821 Occupation =Agricultural, Forestry and Fishery Workers
1.357772 Education Level=Nil
1.266226 Education Level=Middle

Cluster 15

4.208450 Age Group=A50_54
1.781228 Marital Status=Divorced
1.294807 Occupation =Agricultural, Forestry and Fishery Workers

The following are clusters formed from the urban dataset

Cluster 0

2.625439 Age Group=A25_29
1.806904 Marital Status=Single
1.798191Education Level=Post Secondary/Tertiary
1.707356 Occupation=Service and Sales Workers
1.646025 Occupation=Other Occupations
1.581759 Marital Status=Married
1.467713 Occupation=Crafts and Related Trade Workers
1.416950 Education Level=Nil
1.411821 Education Level=Middle
1.316344 Education Level=Primary

Cluster 1

1.978736 Age Group=A55_59
1.977249 Age Group=A60_64
1.674059 Age Group=A15_19
1.622595 Marital Status=Widow(er)
1.423156 Occupation=Agricultural, Forestry and Fishery Workers
1.346345 Occupation=Other Occupations
1.152217 attribute Gender=M

Cluster 2

4.530735 Occupation=Elementary Occupations
2.610521 Occupation=Plant and Machine Operators, and Assemblers
1.146524 Marital Status=Single

Cluster 3

3.504265 Age Group=A30_34
1.898319 Occupation=Service and Sales Workers
1.760951 Marital Status=Married
1.664710 Education Level=Nil
1.607425 Education Level=Middle
1.533219 Occupation=Crafts and Related Trade Workers
1.421016 Education Level=Post Secondary/Tertiary
1.402696 Education Level=Primary
1.364513 Marital Status=Single
1.335039 Occupation=Other Occupations

Cluster 4

1.780435 Age Group=A20_24
1.733425 Occupation=Other Occupations
1.654433 Marital Status=Single
1.168900Education Level=Post Secondary/Tertiary

Cluster 5

2.994031 Marital Status=Cohabiting

Cluster 6

1.941149 Age Group=A45_49
1.694121 Marital Status=Divorced
1.579777 attribute Gender=M
1.573825 Occupation=Agricultural, Forestry and Fishery Workers
1.494127 Marital Status=Widow(er)
1.006830 Education Level=Middle

Cluster 7

1.970187 Age Group=A50_54
1.716384 Marital Status=Widow(er)
1.629724 Marital Status=Divorced
1.393063 Occupation=Agricultural, Forestry and Fishery Workers
1.280986 attribute Gender=M

Cluster 8

2.413646 Age Group=A35_39
1.733521 Occupation=Service and Sales Workers
1.652614 Marital Status=Married
1.536914 Education Level=Nil
1.472968 Education Level=Primary
1.438735 Education Level=Middle
1.433420 Occupation=Agricultural, Forestry and Fishery Workers
1.338372 attribute Gender=M
1.200335Education Level=Post Secondary/Tertiary
1.034466 Occupation=Other Occupations

Cluster 9

1.913149 Age Group=A40_44
1.715672 attribute Gender=M
1.646109 Marital Status=Divorced
1.563689 Occupation=Agricultural, Forestry and Fishery Workers
1.374523 Marital Status=Widow(er)
1.311274 Education Level=Primary
1.238992 Education Level=Middle
1.178791 Occupation=Service and Sales Workers
1.141582 Marital Status=Married
1.134629 Education Level=Nil

A cluster is a collection of elements with similar characteristic. With respect to the results obtained, each cluster represents characteristics of groups or classes of people prone to the pandemic.

Factors appearing together in the same cluster is an indication that such factors have similar affinity for HIV/AIDS infection. However, some of the socio-economic factors are more prone to being infected with the pandemic than others; and this is indicated by the contrast value. The contrast of *Age Group=A30_34*, for instance, is 5.903213 whereas that of *Marital Status=Single* is 1.307975, but they are both in *Cluster 0* of the Rural setting. This means people in *Age Group=A30_34* are more prone to being infected than those with *Marital Status=Single.*

In terms of content, clusters 0, 5, 6, 11 and 13 of the rural setting are very similar to clusters 0, 3, 8 and 9 of the urban setting. This is evidence that, similar factors characterise the pandemic from both settings. On the other hand, clusters 7, 10 and 12 are unique to the rural setting while cluster 2 of the urban setting is unique.

As seen from the list of clusters, some factors appear in more than one cluster. In order to summarise the associations among factors and corresponding clusters, a bipartite graph by generated. The bipartite graph shows the relationship between the clusters and the factors associated with them. Fig. 1 and Fig. 2 are the bipartite graphs generated for urban and rural settings respectively. If a factor belongs to more than one cluster, then it has a relationship with several groups of factors. For instance, marital status=widow (er) in fig.1 is associated with (appears in) clusters 1, 6, 7 and 9. It therefore has relationship with every factor which occurs in those clusters. For example, through cluster 7, it has a relationship with Age Group 50-54 and through cluster 6, it is related to middle education level. Elaborating on this example, it may mean that, widows and widowers who are aged between 50 to 54 are prone to HIV/AIDS infection. It may also mean that, widows and widowers are at similar risk of contracting the disease as people aged between 50 to 54. Similar logic can be used to explain all the other cases of relationships between factors and clusters.

Previous researches generally paid on only identifying the socioeconomic HIV drivers. In research works such as Nagoli *et al. (*2010), and Bogale *et al*. (2009), the common socioeconomic HIV drivers were identified as lists; without further insight regarding how much impact each driver contributes to the persistence of the epidemic. Without such insight, it is unclear as to which driver to tackle first. Knowledge about levels of impact would advise policymakers how to apportion their limited resources appropriately to efficiently curb the epidemic; such that, drivers with higher impacts are apportioned more resources than those with little impact. Another aspect of HIV-drivers addressed in this paper which previous works missed to address is the concept of interactivity of drivers. The contrasts graphs generated in this paper show some interactions between some of the factors; which is an indication of interdependency. It is therefore possible that, eliminating one factor could lead to the elimination of another factor.

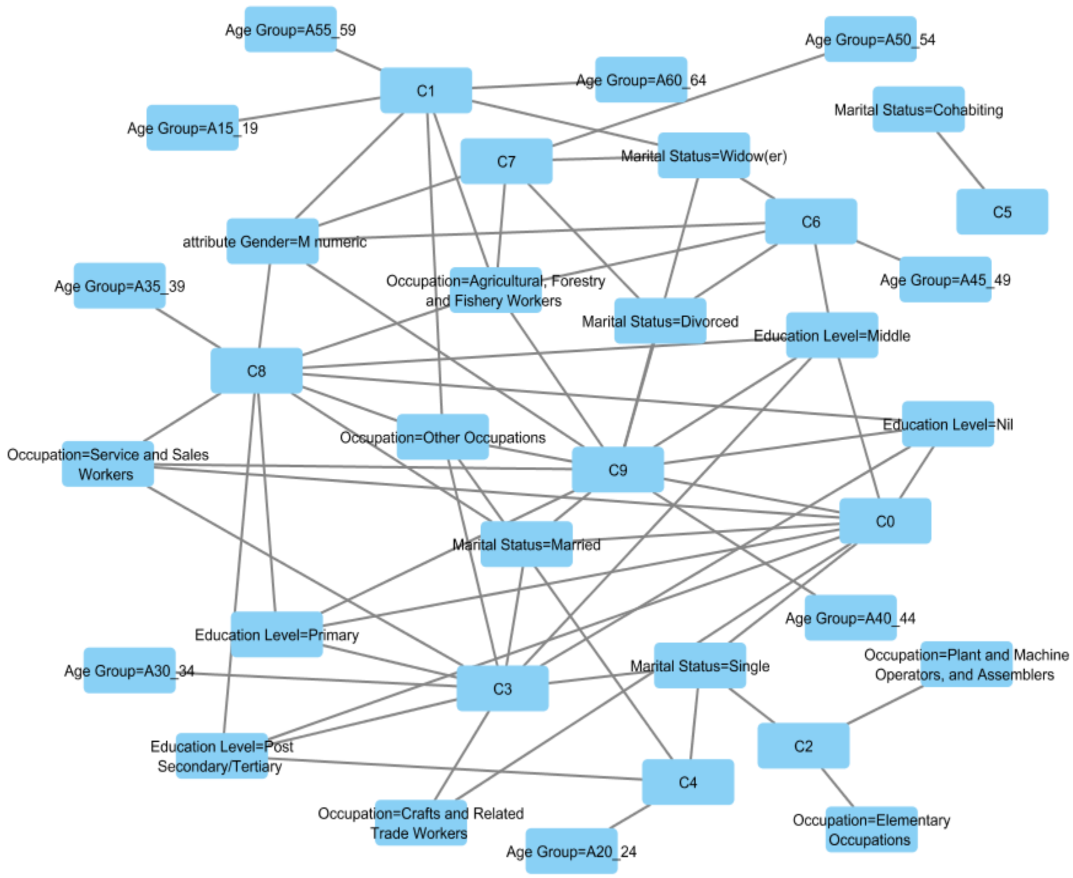

Fig. 1 Graph of Clusters of HIV-Driving factors in rural setting in Ghana

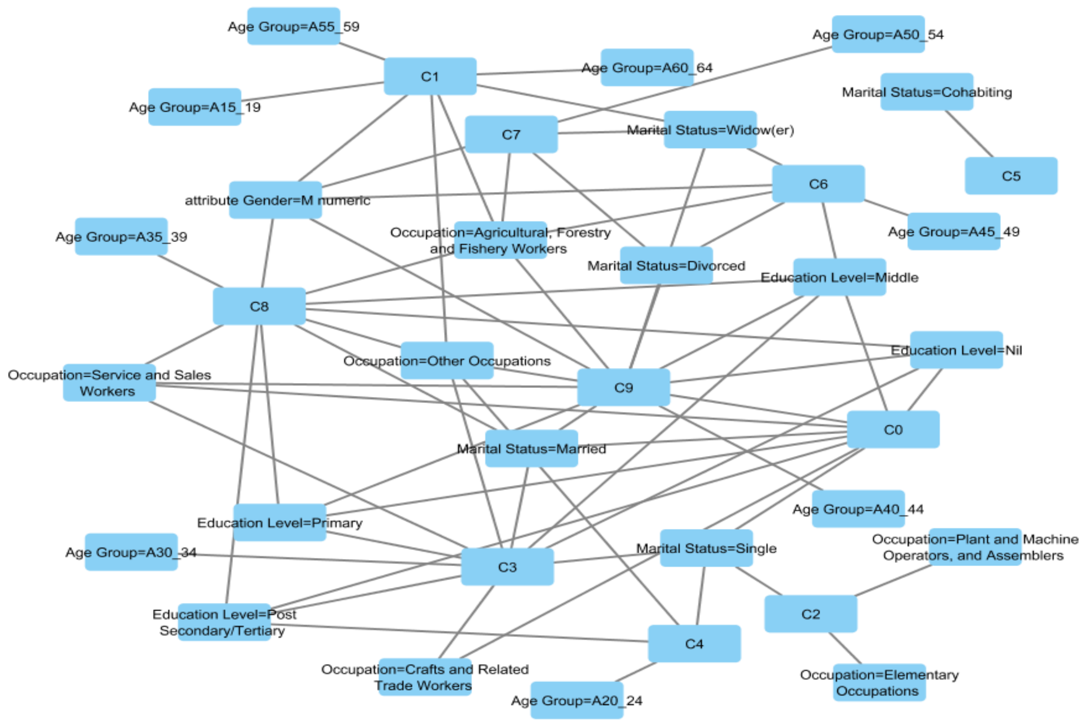

Fig. 2 Graph of Clusters of HIV-Driving factors in in urban setting in Ghana

4 Conclusion

This paper assessed the impact of selected socio-economic drivers of HIV/AIDS that characterised the pandemic in both rural and urban settings in Ghana. It extends earlier research efforts which identified HIV/AIDS socio-economic drivers to measuring the weight of impact. Examining the impact of these drivers is important to help formulate target-specific strategies to curb the disease. The results are useful in guiding researchers and policymakers as to how to direct efforts and resources in fighting the pandemic.

Acknowledgement

The authors would like to express their sincere gratitude to the National AIDS/STI Control Program (Ghana) for making data available for this research.

References

Ankrah, D.N., Lartey, M., Mantel-Teeuwisse, A.K. and Leufkens, H.G., (2017), "Five-year trends in treatment changes in an adult cohort of HIV/AIDS patients in Ghana: a retrospective cohort study", *BMC infectious diseases*, Vol. 17, No.1, pp. 664-674.

Anon. (2012), "International Standard Classification of Occupations 2008 (ISCO-08): Structure, group definitions and correspondence tables", *www.ilo.com*. Accessed: September 11, 2017.

Arrehag, L., Durevall, D. and Sjöblom, M., (2006), "The impact of HIV/AIDS on the economy, livelihoods and poverty of Malawi". *www.sida.se*. Accessed: November 24, 2018

Berkhin, P. (2006), "A survey of clustering data mining techniques". In *Grouping multidimensional data*, Springer, Berlin, pp. 25-71.

Bertozzi, S., Padian, N.S., Wegbreit, J., DeMaria, L.M., Feldman, B., Gayle, H., Gold, J., Grant, R. and Isbell, M.T., (2006), "HIV/AIDS prevention and treatment", Disease control priorities in developing countries, Washington, DC. Oxford University, pp. 331-370.

Biney, E.A., Oduro, G.D., Yar, D.D., Oppong, C.K., Nyame, K., Forson, P.K., Oteng, R., Boakye, I., Norman, B., Ansong, D. and Owusu-Dabo, E., (2015), "HIV/AIDS prevalence at the accident & emergency centre of a tertiary and referral health institution in Ghana", *Ghana Medical Journal*, Vol. 49, No. 4, pp. 220-226.

Blattner, W., Gallo, R.C. and Temin, H.M., (1988), "HIV causes aids", *Science*, Vol. 241, No. 4865, pp. 515-516.

Bogale, G. W., Boer, H. and Seydel, E. R. (2009), "HIV-prevention knowledge among illiterate and low-literate women in rural Amhara, Ethiopia", *African Journal of AIDS Research*, Vol. 8, No. 3, pp. 349–357.

Cohen, M.S., Shaw, G.M., McMichael, A.J. and Haynes, B.F., (2011), "Acute HIV-1 infection", *New England Journal of Medicine*, Vol. 364, No. 20, pp. 1943-1954.

de Walque, D, (2006), "Discordant couples: HIV infection among couples in Burkina Faso, Cameroon, Ghana, Kenya, and Tanzania", *www.openknowledge.worldbank.org*. Accessed: October 4, 2018.

Fortson, J.G., (2008), "The gradient in sub-Saharan Africa: socioeconomic status and HIV/AIDS", *Demography*, Vol. 45, No. 2, pp. 303-322.

Fritzke, B., (1995), "A growing neural gas network learns topologies", *Advances in Neural Information Processing Systems*, MIT press, Cambridge, pp. 625-632.

Heeney, J.L., Dalgleish, A.G. and Weiss, R.A., (2006), "Origins of HIV and the evolution of resistance to AIDS", *Science*, Vol. 313 No. 5786, pp. 462-466.

Hladik, F. and McElrath, M.J., (2008), "Setting the stage: host invasion by HIV", *Nature Reviews Immunology*, Vol. 8, No. 6, pp. 447-457.

Holmström, J., (2002), "Growing Neural Gas--Experiments with GNG--GNG with Utility and Supervised GNG", *Unpublished Master's Thesis*, Department of information technology Computer Systems, Uppsala, Sweden, pp. 1-42

Isiugo-Abanihe, U.C. and Oyediran, K.A., (2004), "Household socioeconomic status and sexual behaviour among Nigerian female youth", *African Population Studies*, Vol. 19, No. 1, pp. 81-98.

Kalichman, S. C., Simbayi, L. C., Kagee, A., Toefy, Y., Jooste, S., Cain, D. and Cherry, C. (2006), "Associations of poverty, substance use, and HIV transmission risk behaviors in three South African communities", *Social Science & Medicine*, Vol. 62, No. 7, pp. 1641-1649.

Kaya, H. O. (2018), "Beyond the Statistics: HIV/AIDS as a socioeconomic epidemic in Africa", In *AIDS and Development in Africa*, Routledge, pp. 37-46.

Lamirel, J. C., and Al Shehabi, S. (2015), "Feature maximization-based clustering quality evaluation: a promising approach", *Trends and Applications in Knowledge Discovery and Data Mining*, pp. 210–222.

Nagoli, J., Holvoet, K. and Remme, M. (2010), "HIV and AIDS vulnerability in fishing communities in Mangochi district, Malawi", *African Journal of AIDS Research*, Vol. 9, No. 1, pp. 71–80.

Ngo, P., Kenmochi, Y., Passat, N. and Talbot, H. (2014), "Topology-preserving conditions for

2D digital images under rigid transformations", *Journal of Mathematical Imaging and Vision*, Vol. 49, No. 2, pp. 418-433.

Parkhurst, J.O., (2010), "Understanding the correlations between wealth, poverty and human immunodeficiency virus infection in African countries", *Bulletin of the World Health Organization*, Vol. 88, pp. 519-526.

Patterson, A.S., (2018), "The Economic, Social, and Political Drivers of the AIDS Epidemic in Swaziland: A Case Study". In *the African State and the AIDS Crisis*, Routledge, pp. 113-142.

Pena, M., Barbakh, W. and Fyfe, C., (2008), "Topology-preserving mappings for data visualisation", In *Principal Manifolds for Data Visualization and Dimension Reduction*, Berlin, Heidelberg. Springer, pp. 131-150.

Saraswathi, S. and Sheela, M.I., (2014), "A comparative study of various clustering algorithms in data mining", *International Journal of Computer Science and Mobile Computing*, Vol. 11, No. 11, pp. 422-428.

Sehgal, G. and Garg, D.K., (2014), "Comparison of various clustering algorithms", *International Journal of Computer Science and Information Technologies*, Vol. 5, No. 3, pp. 3074-3076.

Sharp, P.M. and Hahn, B.H., (2010), "The evolution of HIV-1 and the origin of AIDS", *Philosophical Transactions of the Royal Society Biological Sciences*, Vol. 365, No. 1552, pp. 2487-2494.

Sharp, P.M. and Hahn, B.H., 2011. "Origins of HIV and the AIDS pandemic". *Cold Spring Harbor perspectives in medicine*, Vol. 1, No. 1, pp. a006841

Sisodia, D., Singh, L., Sisodia, S. and Saxena, K. (2012), "Clustering techniques: a brief survey of different clustering algorithms", *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, Vol. 1, No. 3, pp. 82-87.

Authors

William Akotam Agangiba is currently a lecturer at the Department of Computer Science and Engineering in University of Mines and Technology, Tarkwa. He obtained his MSc and BSc degrees in Information Systems and Technologies from the Tver State Technical University, Russia. He is a member of the Institute of Electrical and Electronics Engineering (IEEE) and the Association of Computing Machines (ACM). His research interests include high level programming languages, Data Structures and algorithms, Data Analytics, Web and Mobile Technologies, Expert Systems and Logics of Computer Science.

Millicent Agangiba is currently a lecturer at the Department of Computer Science and Engineering in University of Mines and Technology, Tarkwa. She obtained her MSc from the Tver State Technical University Russia in Computer Complexes, Systems and Networks. She is a member of the Institute of Electrical and Electronics Engineering (IEEE) and the Association of Computing Machines (ACM). She is a Schlumberger Faculty for the Future Fellow and L'Oréal-UNESCO for Women in Science Fellow. Her research interest is in Web and Mobile Technologies, Interaction and Accessibility particularly in the areas of E-government and E-learning, Digital Inclusion for Persons with Disabilities, Information and Communication Technologies for Development.