International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 16 (2024) pp. 040-053

Received: 20 July 2022; Accepted: 28 August 2023; Published: 24 April 2024

Gaussian graphical model for Perceiving the Sense of Community of Melanau Coastal Village in Mukah, Sarawak

Mohamad Hardyman Barawi[1], Noor Fazilla Abd Yusof[2], Donald Stephen[3], Mohd Kamal Othman[4], Zaimuariffudin Shukri Nordin[5] and Mukhriz Izraf Azman Aziz[6]

[1]Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak
94300 Kota Samarahan, Sarawak, Malaysia
bmhardyman@unimas.my

[2]Fakulti Teknologi Maklumat Dan Komunikasi,Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
elle@utem.edu.my

[3]Centre for Foundation and General Studies, i-CATS University College
Jln Stampin Timur, 93350 Kuching, Sarawak, Malaysia *donald@icats.edu.my*

[4]Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak
94300 Kota Samarahan, Sarawak, Malaysia
omkamal@unimas.my

[5]Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak
94300 Kota Samarahan, Sarawak, Malaysia
nzaim@unimas.my

[6]School of Economics, Finance and Banking, Universiti Utara Malaysia
Changlun, 06050 Bukit Kayu Hitam, Kedah, Malaysia
mukhriz@uum.edu.my

***Abstract*: This paper reports the findings of the Sense of Community (SoC) investigation using a partial correlation network analysis model within the Melanau community in Mukah, Sarawak. This study intended to investigate the neighbourhood structure and interactions among Melanau households. A sample of 85 households from three (3) Melanau coastal villages have participated in the survey and completed the questionnaires. The collected data has been analysed using regularised partial correlation network analysis (EBICglasso). The findings from this study proved that the support from leaders and community members are strongly related. However, the connection between these two nodes and other nodes in the network is relatively weak. This study also suggests that caring for other community members is the most central node; the participation in the community events is the most peripheral node in the sense of community network among the Melanau community. The insights from this study shed light on the associations of neighbourhood environment and sense of community. Thus, facilitating an informed decision-making process which could help researchers and policymakers understand the climate of SoC for intervention planning in the Melanau community.**

***Keywords*:** Network Analysis, Sense of Community, Rural communities, Borneo, Coastal Village

I. Introduction

A sense of Community (SoC) is the sense of connectedness or attachment to each other and the group and sharing closeness and affection through commitments [43]. SoC focuses on the individual experience and perception during social interaction with other community members. SoC is a term widely used in community psychology, social psychology, and community social work research. For example, in community psychology, SoC involves community involvement, empowerment, social harmony, personality, and standards of living [13]. In addition, [29] describe SoC through shared emotional connection, influence, reinforcement and membership. However, researchers have widely used other concepts, such as cohesion and attachment, social capital and social support, as a part of the SoC [35]. Furthermore, research has investigated the SoC in several domains, for example, in online and collaborative learning [11, 40], crisis manage-

ment [8], living conditions in rehabilitation facilities [58], and neighbourhood environment [54]. Extant literature, for example, [35, 15, 26] has extensively discussed SoC and its impact on the community building, mental health, individual resilience, social disorders, civic responsibility, safety and security. In addition, extensive research on SoC through sixty decades of study in the United States suggested that having SoC is strongly correlated with lower mental, social and health disorders [12]. Although various studies have focused on SoC and highlighted several critical findings mentioned above, there is a lack of studies on SoC and its impact on the local community in Malaysia. A SoC study in Malaysia has widely focused on the community leaders [51, 36], business-related industries such as homestay operators [49, 37]. Hence this study will focus on the specific rural coastal village in Mukah, Borneo. In addition, different methods have been used to analyse the SoC, for example, qualitative and quantitative methods. However, no known studies utilise the Graphical Gaussian model (GGM), especially in the Malaysian context. Thus, the goal of this study is to report the SoC findings using the GGM.

Formerly, the latent variable model has been widely used to investigate the correlation between the SoC and contex- tual factors. The latent variable model can also be broadly described as a hypothetical variables model that is not visible within the dataset to measure real-world phenomena. The latent variables are used interchangeably in the literature, e.g., constructs, unobserved/unmeasured variables, and factors [6]. Researchers infer these hypothetical unmeasurable/ unobserved variables from the data set of observable variables (or measured directly). In a typical instrument design for standard factor analysis, these hypothetical variables are assumed to "cause" people to respond to the observed variables [6]. Generally, the latent variable model has been widely used in psychology or psychological measurement [60, 44]. [9] used latent variables to explore the effects of one's working memory in updating and integrating information. [38] suggested the correlation among the group norms, multi-sensory styles and mobility patterns with the latent variable analysis. Researchers also used the latent variable interaction to demonstrate the effects of racial discrimination among Asian Americans [4].

In recent studies [60, 33, 46, 42], the performance of network analysis outperformed the latent variables in modelling the psychological construct. In contrast to the latent variable model that summarised several items into a latent variable score (composite score), the network approach does not yield any (composite score) to represent the latent variable. Otherwise, it depends on the interaction between items to illustrate a psychological phenomenon. The application of network analysis has gained increasing popularity in psychology (it was initially introduced as a psychometric framework) to demonstrate and map the psychological constructs [18].

A psychological network is made up of items (represented by circles called nodes) and lines called edges that illus- trate relationships between nodes [19, 20]. To get insights into which variables/items predict one another, the Gaus- sian Graphical Model (GGM) is among the most widely ap- plied. This model is commonly suited for polytomous vari- ables [1, 24]. The GGM is essentially the inverse of all nodes' variance-covariance matrix. When the inverted matrix is standardised, it produces a partial correlation matrix [19, 33]. Partial correlations depict the degree of relationships between nodes while considering the effects of the network's other nodes [33]. These relationships can also be estimated by computing node-wise regression.

Based on the aforementioned studies, the GGM illustrates emerging patterns of sophisticated item-level interactions and conditional dependence relationships. Also, it was not available when using the latent variable approach to psychometrics. Contrarily, unlike the latent variable approach, the items in the SoC instrument are not viewed as "symptoms" of a good SoC but rather a system of an interconnected network of all elements in SoC [32]. Therefore, this study aims to investigate the correlation between SoC and neighbourhood environment within the Melanau community using partial correlation network analysis based on GGM.

II. Related Work

A. Gaussian Graphical Model (GGM)

Gaussian Graphical Models (GGMs) are a robust tool for representing and analyzing multivariate data. The tool can be use to uncover the underlying structure of conditional independence relationships among variables, make accurate predictions, and tackle various tasks like clustering and dimensionality reduction. A GGM is made up of a series of components or variables denoted by nodes and lines (edges) explaining the correlations between the components or variables [20, 7]. The correlations between the relevant components or variables are shown by these lines (edges) wideness, and the lack of lines denotes null or extremely weak correlations between the relevant components or variables. These GGM lines represent partial correlations, or the correlation among two components or variables once all components or variables in the sample group are contained. The GGMs are used to characterise provisional correlations among two or more variables. This goal can be achieved by deciding which precision matrix off-diagonal elements to use. For exam- ple, the covariance matrix's inverse is non-zero [65]. Al- though the visual representation of correlations can help researchers gain quick insights into the data, GGM can sometimes be challenging to interpret if the projected graphs are crowded and consist of many lines. Moreover, absolute zero partial correlations are rare due to sample deviation, resulting in concentrated graphs and containing fallacious relationships [20, 61].

Gaussian Graphical Models (GGMs) encompass various types of models that are used to represent and analyze multivariate data. Here are some common types of GGMs along with their explanations and references:

1) Gaussian Markov Random Fields (GMRFs)

A Markov random field is used to represent the underlying graph structure in GMRFs. The conditional independence relationships among variables in a GMRF are encoded in the precision matrix (inverse covariance matrix), which determines the graph's edges. GMRFs have been studied and applied extensively in a variety of fields, including spatial statistics, image analysis, and Bayesian modelling [55]. The use of GMRFs has a number of advantages. First, they enable efficient modelling and inference in high-dimensional settings with a large number of variables in comparison to the sample size. The sparsity assumption in GMRFs, which assumes that the graph structure is sparse, aids in model estimation and interpretation. The sparsity assumption implies that given the remaining variables, most variables are conditionally independent of each other, resulting in a sparse precision matrix. GMRFs have found use in spatial statistics, where they are used to model spatial dependencies between observations. GMRFs can capture the inherent spatial structure of data and account for spatial autocorrelation by incorporating spatial information. As a result, they are particularly useful in geostatistics, environmental modelling, and disease mapping. In addition, GMRFs have been used in image analysis tasks such as denoising, segmentation, and reconstruction. GMRFs can effectively capture the spatial structure of images and facilitate various image processing tasks by treating pixel intensities as variables and modelling the dependencies between neighbouring pixels.

2) Sparse Gaussian Graphical Models

Sparse GGMs are based on the assumption that the underlying graph is sparse, which means that only a subset of variables are directly connected by edges. This assump- tion enables more efficient estimation and interpretation of graph structure, especially in high-dimensional settings. Sev- eral methods for estimating sparse GGMs have been devel- oped, including penalised maximum likelihood and graphi- cal LASSO [27]. The sparsity assumption in GGMs allows for more efficient graph structure estimation and interpreta- tion. GGMs focus on identifying the most relevant and informative connections between variables by assuming sparsity, reducing model complexity and improving interpretability. This is especially important when the number of variables is large and traditional dense models would be computationally difficult or impractical. Several methods for estimating sparse GGMs have been developed. Penalised maximum likelihood estimation is a popular approach that adds a penalty term to the likelihood function to encourage sparsity in the estimated precision matrix. This penalty term limits the number of non-zero elements in the precision matrix, encouraging the selection of only the most significant edges. Graphical LASSO (Least Absolute Shrinkage and Selection Operator) is another popular method that combines a LASSO penalty with maximum likelihood estimation. The LASSO penalty effectively reduces the precision matrix estimates to zero, encouraging sparsity and variable selection. Graphical LASSO has been demonstrated to perform well in high-dimensional settings and is commonly used to estimate sparse GGMs. These methods have significantly improved our ability to estimate sparse GGMs and identify the underlying graph structure in high-dimensional datasets. Sparse GGMs provide more interpretable and manageable models for analysing complex datasets by focusing on the most relevant connections between variables.

3) Group Sparse Gaussian Graphical Models

Group sparse GGMs are Sparse Gaussian Graphical Mod- els (GGMs) that incorporate group structures or prior knowledge about the variables. These models offer a framework for identifying groups of variables that are conditionally independent of one another. Group sparse GGMs improve our ability to capture and interpret relationships between variables in a variety of fields, including genetics, neuroscience, and social network analysis, by taking into account the group structure [66]. Group sparse GGMs provide a flexible approach to modelling variable dependencies within specific groups or subsets. Variables in many real-world scenarios can naturally be classified into distinct groups based on shared characteristics or functional relationships. Group sparse GGMs enable the discovery of conditional independence relationships within each group by accounting for these group structures, enhancing our understanding of the underlying data. Group sparse GGMs, for example, can be used to analyse gene expression data in genetics, where genes can be organised into groups based on their biological functions or pathways. Researchers can uncover dependencies between genes within each group by incorporating the group structure, which may have important implications for understanding gene regulation and interactions. Group sparse GGMs can be used to study brain connectivity networks in neuroscience. Anatomical regions or functional networks can be used to categorise brain regions or neuronal populations. Group sparse GGMs enable researchers to identify sets of brain regions that exhibit conditional independence by taking into account the group structure, shedding light on the functional organisation of the brain. Another application for group sparse GGMs is social network analysis. Individuals in social networks are frequently grouped into communities based on shared characteristics or social affiliations. Group sparse GGMs can assist in identifying groups of individuals who have independent relationships within their respective communities, allowing for a more in-depth understanding of social dynamics and interactions. The ability of group sparse GGMs to incorporate group structures or prior knowledge about variables makes them an effective tool for modelling complex relationships in a wide range of fields. These models improve our understanding of the underlying data, facilitate interpretation, and provide insights into the underlying mechanisms and interactions by capturing conditional independence within groups.

4) Dynamic Gaussian Graphical Models

Dynamic Gaussian Graphical Models (GGMs) are designed specifically to capture the time-varying dependencies between variables in a time series. These models expand on the concept of GGMs by incorporating temporal dependen- cies, enabling the modelling and analysis of time series data in a variety of fields such as econometrics, finance, and neuroscience [5]. Time series data frequently exhibit tempo- ral dynamics, or changes in the relationships between vari- ables over time. Dynamic GGMs provide an effective frame- work for capturing and representing these changing depen- dencies. By taking into account the temporal dimension, dy- namic GGMs allow for the modelling of how relationships between variables evolve and adapt over time. Dynamic

GGMs are widely used in econometrics and finance to model and analyse financial time series data, such as stock prices or asset returns. These models enable researchers to capture the changing dependencies and correlations between finan- cial instruments, which is important for tasks such as portfo- lio management, risk assessment, and forecasting. Dynamic GGMs provide a more accurate representation of the underlying dynamics of financial markets by incorporating time-varying dependencies. In neuroscience, dynamic GGMs are used to analyse brain connectivity networks over time. Neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) or electroencephalography (EEG), provide measurements of brain activity that change over time. Dynamic GGMs can capture changing patterns of connectiv- ity between brain regions, allowing researchers to study the dynamics of brain networks and investigate how they relate to cognitive processes or neurological conditions. Dynamic GGMs are used in fields other than econometrics, finance, and neuroscience. They've also been used in fields like environmental monitoring, engineering, and social sciences, where time series data with evolving dependencies are common. State-space models, time-varying graphical LASSO, and Bayesian approaches can all be used to estimate dynamic GGMs. These methods can estimate the time-varying precision matrix or its inverse covariance matrix, which captures the changing relationships between variables over time.

B. Estimating the Gaussian Graphical Model (GGM)

One common method for estimating a structure with GGMs in high dimensions is based on the classical re- sult that the zeros of a precision matrix converge to zero partial correlation, which is both appropriate and neces- sary for conditional probability [41]. Assuming only a few conditional dependencies corresponds to imposing a sparsity constraint on the precision matrix entries, result- ing in a combinatorial problem. Some of the popular ap- proaches to learning GGMs can be viewed as using the ℓ to create convex surrogates to this problem. Furthermore, a combinatorial problem results from several conditional dependencies assumptions that match a sparsity (parameter) limitation on the precision matrix entries. Some of the common methods for determining GGMs could be viewed as using the ℓ to create convex surrogates to this problem. As a result, the glasso algorithm is a frequently used approach in GGM to achieve a sparser graph [27, 10, 52, 59].

However, this approach requires regulators to use a tuning parameter that forces the small partial correlation coefficients to zero, resulting in sparsity. Thus, the different graphs result from different tuning parameter values. For example, several low tuning parameter values will result in a dense graph, whereas several high tuning parameter values will produce sparse graphs. The greatest interactions maintained in the graph (those which maximise true positive) are used to determine the optional tuning parameter setting, as shown in the extended Bayes information criteria (EBIC) [25].

The technical aspects of the GGM are beyond the scope of this paper; hence, it is suggested that the readers read the work by [24] to understand the measurement of GGM models and their application focusing in psychology.

C. GGM vs Other Models

Theoretically, GGM is similar to other exploratory modelling methods in psychology, specifically exploratory factor analysis (EFA), which is used to investigate the correlation between study components. Furthermore, the GGM and unidimensional factor models have similarities at the component level [39, 62, 64, 19]. The measured variables in a unidimensional factor model are uncorrelated to the latent variable, also known as a one-factor model. This implies that after considering the latent variable, the correlations between components should tend to be zero. As a result, a cluster of items that are fully connected suggests that these components are measuring the same latent variable.

As a result, at the component level, such a similarity can also be used to understand the questionnaire's factor structure [56, 31, 17]. However, the GGM vary from standard exploratory data exploration subject to partial correlational coefficients. Evidently, a GGM depicts correlations between components and variables as a graph that is easier to comprehend compared to a big partial correlation table, especially if small correlation coefficients have become zero using the glasso algorithm, as demonstrated in this study.

D. Application of the GGM

In this study, we demonstrate the applicability and importance of the GGM for environmental psychologists and other related studies in investigating relationships between components and variables from three Melanau coastal fishing villages.

Our research aims to provide nuanced insights into the factors underlying the Melanau people's lives while also providing a compelling lens through which environmental psychologists can gain a profound understanding of the intricate interplay between individuals and their environments by focusing on a small but meaningful dataset.

Specifically, the study aims to shed light on the associations of neighbourhood environment and SoC in three coastal fishing villages. Hence, we argued that neighbourhood environment aspects are important in discerning the SoC, which could help policymakers understand the climate of SoC for intervention planning in the Melanau community.

III. Materials and Methods

A. Background of Location

Mukah (refer Figure 1[1]) is one of the districts in Sarawak, Malaysia which has an area of 2,536 sq km^2. Initially, Mukah town was opened by Melanau indigenous group before it was governed by the British explorer, James Brooke between 1841 and 1946. Mukah town is a small coastal town to support the government administration within the Mukah division. The Mukah division consists of five (5) districts, i.e., Mukah Town, Dalat, Matu, Daro and Tanjung Manis. The town of Mukah was established on the edge of the Batang Mukah river, and it is about a 2.5 hour's drive from Sibu, Sarawak. It has a long coastal leading to the Bintulu regency.

[1]https://kiswa.sarawak.gov.my/kiswa/

[2]https://mukah.sarawak.gov.my/page-0-0-349-Latar-Belakang.html

The position of Mukah facing the South China Sea causes Mukah to face the monsoon season during the windy season when the Northeast monsoon is blowing. Due to the direct connection of the Mukah river and Balingian river to the sea, it has become the strategic route for entry and exit from the city of Mukah (for fishing boats and oil tankers).

According to the Sarawak population by District 2017 census[3], the total population of Mukah Town is 49,900 with 20,900 being dominated by ethnic Melanau, 13,700 Iban, 3,900 Chinese, 3,300 Malay, and 200 others. On average, the Melanau community inhabits a total of 36 villages in the Mukah area. The villages in the Mukah district are located on the edge of the main river i.e., the Mukah river and its areas include the Tellian river, Penakub river, Petaanak river and Penipah river. The Melanau tribe mostly inhabits villages throughout the Mukah area compared to only a few from other tribes such as tribes Chinese and Iban.

The villages in Mukah are usually inhabited separately, following the beliefs of the Melanau community. In general, there is a separation from the placement of the residents of Mukah. For example, the Muslim Melanau dominates the placement of villages around Mukah City and areas close to rivers, such as Kampung Litong, Kampung Alo River, Kuala Lama Village and Penakub Ulu Village, whilst the Christian community, they dominate the village far from city centres, such as Tellian Daya Village, Petanak Daya Village, Jebungan Village and Teh Village.

B. Sample selection procedure during MCO

The study was conducted during the period of nationwide lockdown or commonly referred to as MCO in response to the COVID-19 pandemic in the country from March 2020 to May 2020. Due to the restrictions imposed by the government, researchers faced obstacles to administer the questionnaires to the respondents. From the minimum 225 samples required, the study only managed to interview 85 household heads from the Melanau coastal villages in Mukah, Sarawak of which two were discarded from further analysis, due to inconsistency and deemed not representative of the population. Therefore, we retained 83 household heads for analysis in this study which corresponds to the households from Tellian Laut village (n = 48), Tellian Tengah village (n = 12), Tellian Ulu village (n = 22). The head of household rated their agreement on each item based on a Visual Analog Scale (VAS) score ranging from 0 (lowest) to 10 (highest). Trained enumerators were assigned to assist the respondents and provide explanations as needed. Missing values in the data were treated using a random forest missing data imputation algorithm implemented in R in the *MissForest*() package [57]. Studies [2] have shown that *MissForest*() handles missing values and produces the lowest imputation error compared to other imputation algorithms. Due to the small-sample size limitation in this study, we utilised the Gaussian Graphical Model (GGM) method to overcome the challenges as it is an efficient model for handling low-sample sizes. sizes [63].

C. Data Preparation and Analysis

A statistical software R version 3.4.1 [50] to employed the Gaussian graphical model. was used to analyse the GGM in this study. We first adopted an imputation method (missForest), an approach using a random forest imputation algorithm to handle missing values in data following [57] work. The approach is implemented using the *missForest*() package in R [53].

In this approach, the missForest first uses the mean to impute all missing data. Subsequently, to calculate the missing values for each variable, it integrates a random forest for each observed section and predicts the missing section. Such a process of training and prediction is conducted iter- atively based on the predefined number of iterations or after the stopping criterion is reached. We used the EBICglasso (Extended Bayesian Information Criterion Graphical Least Absolute Shrinkage Operator) function regularisation technique [19] in computing the Gaussian graphical model to estimate the sense of community (SoC) partial correlation network.

Obtaining partial correlation network While there are other methods for calculating partial correlation coefficients [14], we will focus on one of the regularly used methods that have been proved to produce swift partial correlations.

To begin, partial correlations can be computed directly from the inverse of a variance-covariance matrix. Let y represent a set of item responses that we may assume to be centred without losing generality. Let Σ represent a variance-covariance matrix. Then, we suppose that y has a multivariate normal distribution as follows:

$$y \sim (0, \Sigma) \tag{1}$$

Let K (kappa) denote the inverse of Σ, also termed the *precision matrix*:

$$K = \Sigma^{-1} \tag{2}$$

then, element k_{ij} (row i, column j of K) can be standardised to acquire the partial correlation coefficient between variable y_i and variable y_j , after conditioning on all other variables in y, $y_{-(i,j)}$ [34]:

$$Corr(y_i/y_{-(i,j)}) = \frac{-k_{ij}}{\sqrt{k_{ij}}\sqrt{k_{ij}}} \tag{3}$$

Interpreting partial correlation networks Partial correlation networks enable a variety of powerful inferences. These summaries are derived from a broader and more detailed technical introduction by [23].

- Partial correlation networks enable the modelling of one-of-a-kind interactions between variables. If A correlates with B and B corresponds with C, we would expect A to correlate with C as well. An unconditional correlation of zero between A and C is unusual since only a few causal structures might result in such a correlational pattern[4]. If the data is normal, partial corre-

[3] https://data.sarawak.gov.my

[4] There are two possibilities: B is a shared effect of A and C, or two orthogonal latent variables generate covariation between A and B and B and C.

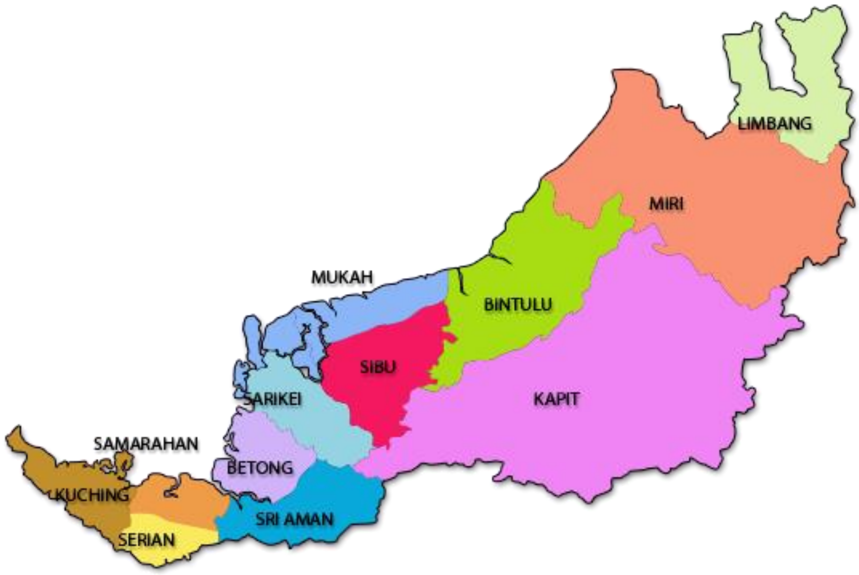

Figure. 1: Map of Sarawak, Malaysia

lations can be read as pairwise interactions[5], of which we only require two to explain the correlational pattern: one between A and B and one between B and C. This model will have one degree of freedom, resulting in a testable hypothesis [22].

- Partial correlations might point to plausible causal pathways. Conditional independence relationships, such those reflected by partial correlation coefficients, are important in causal inference [16]. When all relevant variables are assumed to be observed (i.e., no latent variables), a partial correlation between variables A and B is only expected to be nonzero if A causes B, B causes A, A and B have a reciprocal relationship, or both A and B cause a third variable in the network [48]. To that purpose, partial correlation networks are regarded as highly exploratory hypothesis-generating structures that indicate probable causal effects. While there are exploratory algorithms for discovering directed (causal) networks, they are based on strong assumptions such as acyclity (a variable may not eventually cause itself e.g., $A \implies B \implies C \implies A$), and are more strongly influenced by latent factors that cause covariation (latent variables produce directed edges between observable variables, reflecting a strong causal hypothesis). Furthermore, these models are difficult to identify and parameterize: Many distinct parameterized equivalent directed models can fit the data equally well. In contrast, partial correlation networks are accurately identified (no equivalent models) and simply parameterized using partial correlation coefficients. As a result, exploratively estimating undirected networks is a more appealing option than exploratively estimating directed networks since it avoids the bothersome and poorly recognised direction of effect[6].
- Latent variables may be highlighted through network clusters. While partial correlations seek to highlight unique variance between two variables, they also preserve common variance owing to outside causes that cannot be completely eliminated by controlling for other variables in the network. As a result, if a hidden variable generates covariation between two or more network variables, all of these variables are anticipated to be linked in the network to form a cluster [30]. As a result, such clusters may be indicative of hidden variables [22].

After estimating the partial correlation networks, less salient relationships (or spurious relations) between items are optimised using an optimal level of LASSO hyperparameter γ [33]. As a result, small edge weights are reduced to be precisely zero. This step increases the saliency of prominent edges in the network and thus eases interpretation.

In this study, we fixed the LASSO hyperparameter γ tuning to 0.5 to ensure that the network does not mask prominent edges nor reveal non-salient (spurious) edges that are considered a nuisance. We then estimated the centrality indices and examined the accuracy of edge-weights by mapping bootstrapped Cis using the bootstrap network estimation methods *bootnet*() package in R [19, 21].). In addition, we validated the perceived community safety using the Andrich rating scale model [3, 1].

IV. Results and Discussion

This section shows the estimated network based on nine nodes of the perceived SoC. Details of each node in the network as illustrated in in Table 1, i.e., Leader, Community, Care, Festives, Participate, Proud, Active, Happy, and Helping. We show the correlation heatmap of the perceived sense of community in Figure 2. The correlation heatmap shows that all nine nodes are positively correlated, with $r > 0.04.$,

[5]This is not the same as the interaction effects of two variables on an outcome variable.

[6]Conditioning on a shared effect can generate an edge in the partial correlation network, therefore it should not be understood as the skeleton of a causal model (a directed network with arrowheads removed). Furthermore, latent variables can generate edges in both directed and undirected networks.

and it indicates that all node pairs are positively correlated without any confounding effects.

Node	Description
Ldr	Leader (support from leaders)
Cmm	Community (support from other community members)
Cmm	Community (support from other community members)
Car	Care (caring about others)
Fst	Festives (joining festivities)
Prt	Participate (Participate in community events)
Prd	Proud (feeling of pride being a member of the community)
Act	Active (active member of the community)
Hpp	Happy (level of contentment)
Hlp	Helping (willingness to help others in the community)

***Table 1*:** Description of nodes in Partial Correlation Network for Sense of Community SoC

We then used the Fruchterman-Reingold algorithm [28] to determine the layout of the Pairwise Markov Random Field (PMRF) network structure. We set the threshold to a positive value to increase the specificity of the network (inadvertently decreases the sensitivity). As a result, the edges of negative partial relationships are becoming more salient (refer Figure 3).

As shown in Figure 3, there are nine nodes in the net- work. The edges (lines) are colour-coded to show the po- larity among nodes in the network. The blue edges denote positive relationships, whereas the red edges denote negative relationships. The Leader (Ldr) and the Community (Cmm) are the most prominent positive relationships. Our result is consistent with the findings from [51], who showed that effective community leaders were able to foster the development of their community by acting as a catalyst for it. On the contrary, the Care (Car) and Happy (Hpp) depicted a negative correlation. This result compliment with the earlier finding of [47], who document that the need for caring for others were negatively related to loneliness and depression and positively related to self-esteem among Korean college students.

The estimated partial correlations network in Figure 3 shows the possible mediation pathways, i.e., Festives (Fst) and Happy (Hpp) are not directly connected as there is no link between them. However, Festives (Fst) affects Active (Act), which in turn affects Happy (Hpp). Therefore, this pathway serves as evidence for the mediation role of Active (Act) in the relationship between Festives (Fst) and Happy (Hpp). It is also noted that although Leader (Ldr) and Community (Cmm) are strongly related, the connection between these two nodes and other nodes in the network is rather weak. Another node with a weak connection is Participate (Prt) due to losing connection to two other nodes.

We also interpret the network structure by examining the Shortest Path Length (SPL) between nodes (Refer Table 2). The existence of intermediary nodes (nodes in between two nodes of interest) increases the path length between these two nodes because the time taken for two nodes to interact increases as the number of intermediary nodes grows [45]. We observed that the Care (Car) has the lowest range of SPL values (0.000 – 6.590), evidence that Care (Car) could be an integral node in the network.

In addition to assessing the network's connectivity patterns, we used centrality indicators such as degree, closeness, betweenness, and predicted influence to determine the network's influential nodes. These centrality indices provide useful information about the relative importance of individual nodes inside the network. Table 3 presents the estimated centrality measures, shedding light on the network's influential nodes. Peripheral nodes, which have low centrality, have fewer connections to other nodes and hence have less influence on overall network dynamics. Central nodes, on the other hand, have a high centrality, showing their criti- cal importance as network components. These nodes have more connections, indicating stronger linkages and higher influence within the network structure. We acquire a better knowledge of the network's dynamics and discover crucial nodes that play important roles in determining information flow, interactions, and overall network cohesion by analysing centrality measurements. Such insights allow for a more nuanced assessment of the network's structure, as well as the identification of nodes that may be particularly crucial in influencing network behaviour and consequences. We plotted the centrality indices of all nodes in Figure 4. We observed that the most central node is Care (Car), and the most peripheral node is Participate (Prt). We then isolate these two nodes and compare the differences (refer to Figure 5). As depicted in Figure 5, the most central node (Care (Car)) shows no direct connection to three other nodes (Cmm, Prt, Act) only, as compared to six nodes for the least central node (Participate (Prt)).

After estimating network structure and centrality indices, network accuracy is estimated. For this purpose, the network edge (line) weights are bootstrapped ($n = 1000$). Figure 6 shows the non-parametric bootstrapped confidence intervals of the estimated edge-weights. The red line denotes the sample values, and the grey area denotes the bootstrapped confidence intervals. The Y-axis represents edge pairs in descending order based on the edge weight, while X-axis represents the bootstrapped confidence interval (CI). We have removed labels on the Y-axis to avoid cluttering and enhance the graph's visibility. As in Figure 6, some edges record wide edge weight CIs, indicating that the network should be interpreted with caution. We noticed that using a small sample size could compromise the edge accuracy.

V. Conclusion

The findings of this study demonstrate the potential mediation effects among the components within the Sense of Community (SoC) through the application of Network Analy- sis. Each identified mediation pathway serves as a hypoth- esis, paving the way for further investigation. It is essen- tial to acknowledge that, akin to other statistical models in psychology, the PMRF model employed in this study simplifies a multifaceted real-world psychological phenomenon. Nonetheless, it offers a valuable and intuitive visualization of the interplay between different components in the Sense of Community.

Moreover, it is important to recognize that psychological research often faces limitations in terms of sample size. In order to enhance the stability and accuracy of the network model, it is recommended to collect a larger number of samples, although this may not always be feasible. Nevertheless,

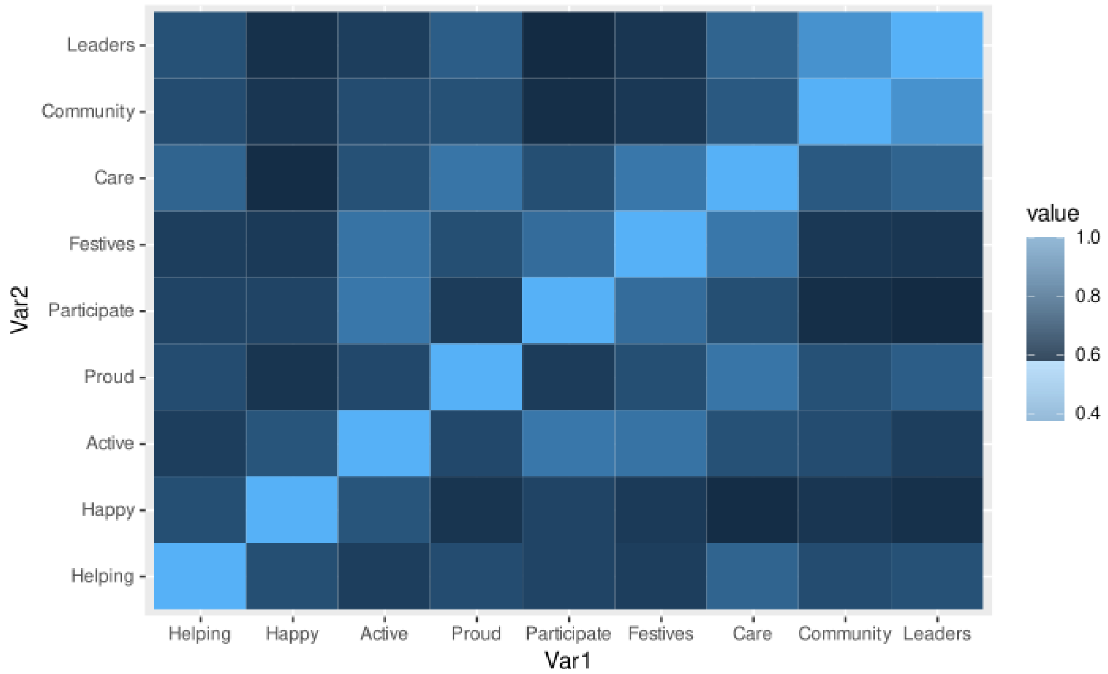

Figure. 2: Correlation Heatmap of perceived Sense of Community

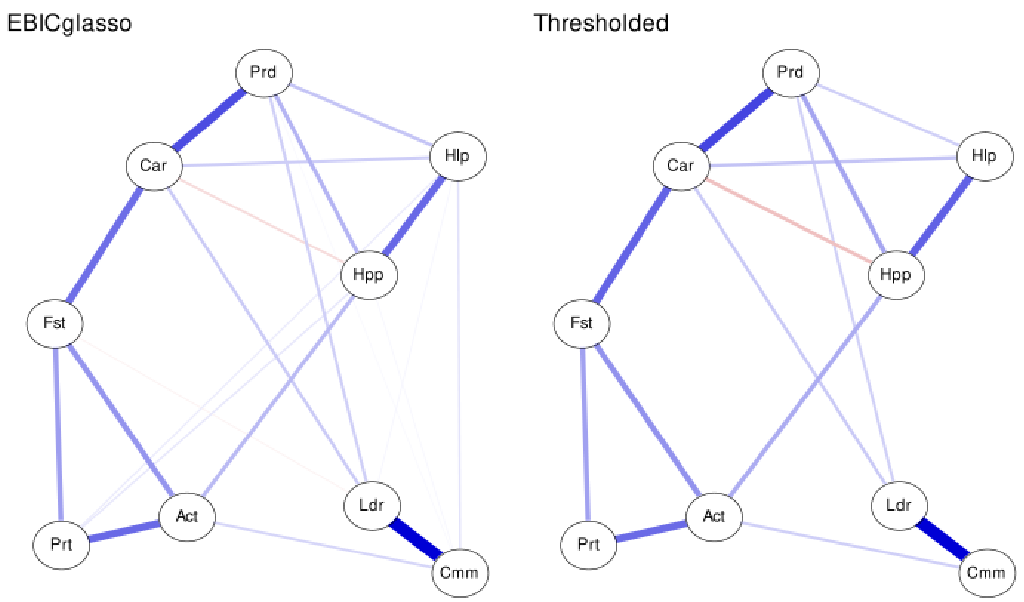

Figure. 3: Partial Correlation Network Structure estimated on Melanau subjects for Perceived Sense of Community

	Hlp	Hpp	Act	Prd	Prt	Fst	Car	Cmm	Ldr
Hlp	0.000	2.260	6.580	6.260	8.980	8.220	5.930	13.920	12.520
Hpp	2.260	0.000	4.320	3.990	6.720	7.590	5.640	11.800	11.600
Act	6.580	4.320	0.000	7.470	2.400	3.270	5.560	7.490	8.890
Prd	6.260	3.990	7.470	0.000	8.010	4.200	1.910	9.000	7.600
Prt	8.980	6.720	2.400	8.010	0.000	3.810	6.100	9.890	11.300
Fst	8.220	7.590	3.270	4.200	3.810	0.000	2.290	10.280	8.870
Car	5.930	5.640	5.560	1.910	6.100	2.290	0.000	7.990	6.590
Cmm	13.920	11.800	7.490	9.000	9.890	10.280	7.990	0.000	1.400
Ldr	12.520	11.600	8.890	7.600	11.300	8.870	6.590	1.400	0.000

***Table 2*:** Shortest Path Length (SPL) between Nodes

	Hlp	Hpp	Act	Prd	Prt	Fst	Car	Cmm	Ldr
CiD	0.746	1.101	1.087	1.041	0.678	1.005	1.458	0.846	0.996
CiC	0.015	0.019	0.022	0.021	0.017	0.021	0.024	0.014	0.015
CiB	0.000	6.000	12.000	2.000	0.000	8.000	16.000	4.000	8.000
CiE	0.746	0.747	1.087	1.041	0.678	1.005	1.104	0.846	0.996

***Table 3*:** Centrality Indices: Degree(CiD), Closeness (CiC), Betweenness (CiB), and Expected Influence (CiE)

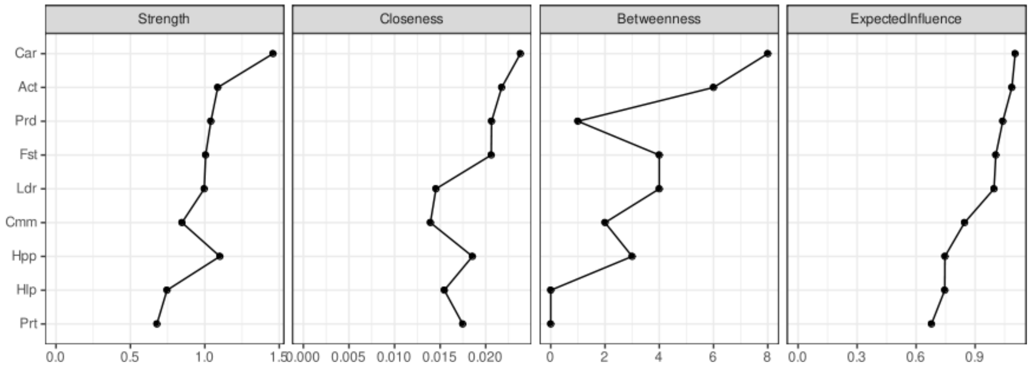

Figure. 4: Centrality Indices of all nodes

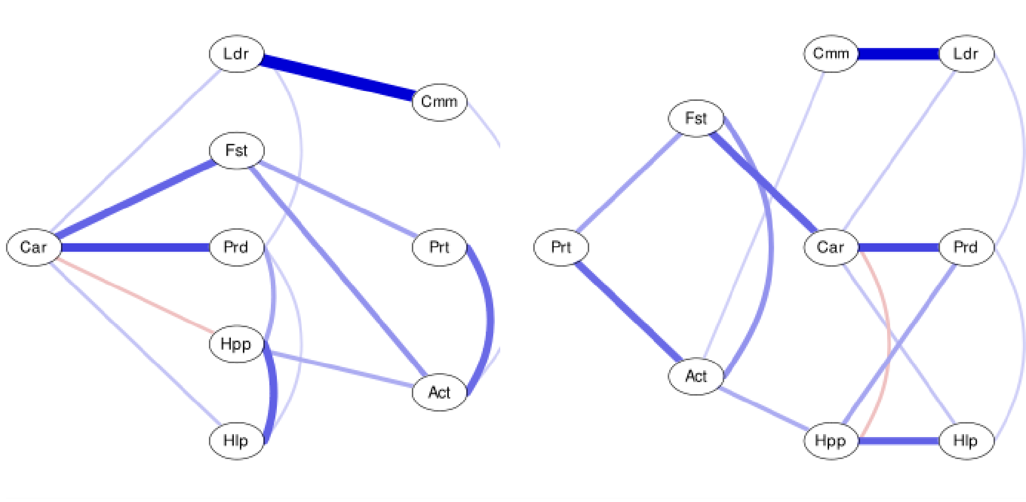

Figure. 5: Network structure difference between the most central & most peripheral nodes

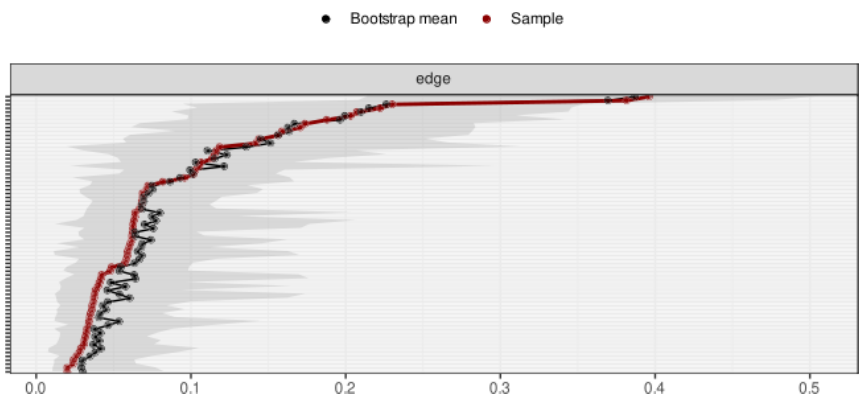

Figure. 6: Bootstrapped confidence intervals (CIs) of the estimated edge-weights

it is imperative for future studies to report the accuracy of the estimated network in order to ensure transparency and reliability.

This study investigates the SoC using a partial correlation network analysis model among the Melanau community in Mukah, Sarawak. The study aimed to examine the neighbourhood structure and the nature of interactions among Melanau households. A sample of 85 households from three (3) Melanau coastal villages have participated in the survey and completed the questionnaires. The collected data has been analysed using regularised partial correlation network analysis (EBICglasso). The findings from this study proved that the support from leaders and community members is strongly related. However, the connection between these two nodes and other nodes in the network is relatively weak. The study also suggests that caring for other community members is the most central node; the participation in the community events is the most peripheral node in the sense of community network among the Melanau community. The insights from this study shed light on the associations of neighbourhood environment and sense of community. Thus, facilitating an informed decision-making process could help researchers and policymakers understand the climate of SoC for intervention planning in the Melanau community.

Acknowledgments

The authors acknowledge Nusantara Chair grant F04/NRC/1904/2019 and Universiti Malaysia Sarawak for supporting this project. The authors would like to thank the reviewers for all of their careful, constructive and insightful comments in relation to this work.

References

[1] Adruce, S. A. Z., Stephen, D., Suhaili, D. S. A. A., John, A. N., Othman, A. M. H. A., and Latif, N. A. (2017). Validation of Perceived Community Safety Instrument Using Polytomous Rasch Measurement Model. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 9(3-11):125–129.

[2] Alsaber, A. R., Pan, J., and Al-Hurban, A. (2021). Handling complex missing data using random forest approach for an air quality monitoring dataset: a case study of Kuwait environmental data (2012 to 2018). *International Journal of Environmental Research and Public Health*, 18(3):1333.

[3] Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4):561–573.

[4] Atkin, A. L., Yoo, H. C., and Yeh, C. J. (2019). What types of racial messages protect Asian American adolescents from discrimination? A latent interaction model. *Journal of counseling psychology*, 66(2):247.

[5] Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008). Model selection through sparse maximum like- lihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516.

[6] Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual review of psychology*, 53(1):605–634.

[7] Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., et al. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, 1(1):1–18.

[8] Boyd, N. M. and Martin, E. C. (2020). Sense of community responsibility at the forefront of crisis management. *Administrative Theory & Praxis*, pages 1–13.

[9] Brydges, C. R., Gignac, G. E., and Ecker, U. K. H. (2018). Working memory capacity, short-term memory capacity, and the continued influence effect: A latent-variable analysis. *Intelligence*, 69:117–122.

[10] Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ 1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

[11] Chatterjee, R. and Correia, A.-P. (2020). Online students' attitudes toward collaborative learning and sense of community. *American Journal of Distance Education*, 34(1):53–68.

[12] Chavis, D. M. and Newbrough, J. R. (1986). The meaning of" community" in community psychology. *Journal of community Psychology*.

[13] Chipuer, H. M. and Pretty, G. M. (1999). A review of the sense of community index: Current uses, factor structure, reliability, and further development. *Journal of Community psychology*, 27(6):643–658.

[14] Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

[15] Colombo, M., Mosso, C., and De Piccoli, N. (2001). Sense of community and participation in urban contexts. *Journal of Community & Applied Social Psychology*, 11(6):457–464.

[16] Constantinou, P. (2015). Extended conditional independence and applications in causal inference. *arXiv: Statistics Theory*.

[17] Crutzen, R. and Peters, G.-J. Y. (2017). Scale qual- ity: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health psychology review*, 11(3):242–247.

[18] Dalege, J., Borsboom, D., van Harreveld, F., and van der Maas, H. L. J. (2017). Network analysis on attitudes: A brief tutorial. *Social psychological and personality science*, 8(5):528–537.

[19] Epskamp, S., Borsboom, D., and Fried, E. I. (2018a). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior research methods*, 50(1):195–212.

[20] Epskamp, S. and Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological methods*, 23(4):617.

[21] Epskamp, S. and Fried, E. I. (2020). Package 'bootnet'. *R package version*, 1.

[22] Epskamp, S., Rhemtulla, M., and Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, 82:904–927.

[23] Epskamp, S., Waldorp, L. J., Mõttus, R., and Borsboom, D. (2016). Discovering psychological dynamics in time-series data. *ArXiv Preprint ID*, 1609(04156):1–31.

[24] Epskamp, S., Waldorp, L. J., Mõttus, R., and Borsboom, D. (2018b). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate behavioral research*, 53(4):453–480.

[25] Foygel, R. and Drton, M. (2010). Extended bayesian information criteria for gaussian graphical models. *Advances in neural information processing systems*, 23.

[26] Francis, J., Giles-Corti, B., Wood, L., and Knuiman, M. (2012). Creating sense of community: The role of public space. *Journal of environmental psychology*, 32(4):401–409.

[27] Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

[28] Fruchterman, T. M. J. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164.

[29] Garrett, L. E., Spreitzer, G. M., and Bacevice, P. A. (2017). Co-constructing a sense of community at work: The emergence of community in coworking spaces. *Organization studies*, 38(6):821–842.

[30] Golino, H. F. and Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS one*, 12(6):e0174035.

[31] Green, S. B. and Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74(1):121–135.

[32] Guyon, H., Falissard, B., and Kop, J.-L. (2017). Modeling psychological attributes in psychology–an epistemological discussion: network analysis vs. latent variables. *Frontiers in psychology*, 8:798.

[33] Hevey, D. (2018). Network analysis: a brief overview and tutorial. *Health Psychology and Behavioral Medicine*, 6(1):301–328.

[34] Højsgaard, S., Edwards, D., and Lauritzen, S. (2012). *Graphical models with R*. Springer Science & Business Media.

[35] Hyde, M. and Chavis, D. (2008). Sense of commu- nity and community building. In *Handbook of commu- nity movements and local organizations*, pages 179–192. Springer.

[36] Islam, M. R., Abd Wahab, H., and ak Anggum, L. (2020). The influence of leadership quality towards community cohesion in iban community in malaysia. *Heliyon*, 6(2):e03370.

[37] Kayat, K., Zainuddin, N. F. A., Ramli, R., and Kasim, M. M. (2016). The perceived role of leadership and community support in the performance of community-based rural homestay (cbrh) programme in malaysia. *International Review of Management and Marketing*, 6(7):145–149.

[38] Krueger, R., Vij, A., and Rashidi, T. H. (2018). Normative beliefs and modality styles: a latent class and latent variable model of travel behaviour. *Transportation*, 45(3):789–825.

[39] Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

[40] Lin, X. and Gao, L. (2020). Students' Sense of Community and Perspectives of Taking Synchronous and Asynchronous Online Courses. *Asian Journal of Distance Education*, 15(1):169–179.

[41] Maathuis, M., Drton, M., Lauritzen, S., and Wainwright, M. (2018). *Handbook of graphical models*. CRC Press.

[42] McFarland, D. (2020). The effects of using partial or uncorrected correlation matrices when comparing network and latent variable models. *Journal of Intelligence*, 8(1):7.

[43] McMillan, D. W. and Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of community psychology*, 14(1):6–23.

[44] McNeish, D., An, J., and Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of personality assessment*, 100(1):43–52.

[45] Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, 32(3):245–251.

[46] Parent, Mike C and Bradstreet, Tyler C and Harmon, Kevin A and McAndrew, Jay and Comiskey, Allison and Cook, A. C. (2018). The Psychology of Men and Masculinities: Using citation network analysis to understand research domains, collaborations, and grant competitiveness. *Psychology of Men & Masculinity*, 19(4):512.

[47] Park, N. and Lee, H. (2012). Social implications of smartphone use: Korean college students' smartphone use and psychological well-being. *Cyberpsychology, Behavior, and Social Networking*, 15(9):491–497.

[48] Parsons, S. (2011). Probabilistic graphical models: principles and techniques by daphne koller and nir friedman, mit press, 1231 pp., $95.00, isbn0-262-01319-3. The Knowledge Engineering Review, 26(2):237--238$.

[49] Pusiran, A. K. and Xiao, H. (2013). Challenges and community development: A case study of homestay in malaysia. *Asian Social Science*, 9(5):1.

[50] R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[51] Rami, A. M., Aziz, F., Zaremohzzabieh, Z., and Ibrahim, A. (2021). Assessing the challenges of local leaders in rural community development: A qualitative study in malaysia. *Pertanika Journal of Social Sciences & Humanities*, 29.

[52] Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980.

[53] Revelle, W. and Revelle, M. W. (2015). Package ‘psych’. *The comprehensive R archive network*, 337:338.

[54] Ross, A. and Searle, M. (2019). A conceptual model of leisure time physical activity, neighborhood environment, and sense of community. *Environment and Behavior*, 51(6):749–781.

[55] Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

[56] Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, 8(4):350.

[57] Stekhoven, D. J. and Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118.

[58] Stevens, E., Guerrero, M., Green, A., and Jason, L. A. (2018). Relationship of hope, sense of community, and quality of life. *Journal of community psychology*, 46(5):567–574.

[59] Tugnait, J. K. (2021). Sparse-group lasso for graph learning from multi-attribute data. *IEEE Transactions on Signal Processing*, 69:1771–1786.

[60] van Bork, R., Rhemtulla, M., Waldorp, L. J., Kruis, J., Rezvanifar, S., and Borsboom, D. (2021). Latent variable models and networks: Statistical equivalence and testability. *Multivariate behavioral research*, 56(2):175–198.

[61] von Klipstein, L., Borsboom, D., and Arntz, A. (2021). The exploratory value of cross-sectional partial correla- tion networks: Predicting relationships between change trajectories in borderline personality disorder. *PloS one*, 16(7):e0254496.

[62] Wainwright, M. J., Jordan, M. I., et al. (2008). Graphi- cal models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305.

[63] Westgate, P. M. and Burchett, W. W. (2017). On the analysis of very small samples of gaussian repeated measurements: an alternative approach. *Statistics in medicine*, 36(6):958–970.

[64] Whittaker, J. (2009). *Graphical models in applied multivariate statistics*. Wiley Publishing.

[65] Williams, D. R. and Mulder, J. (2020). Bayesian hypothesis testing for gaussian graphical models: Conditional independence and order constraints. *Journal of Mathematical Psychology*, 99:102441.

[66] Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.

Author Biographies

First Author Mohamad Hardyman Barawi is currently an academician at the UNIMAS. He received his PhD in Computing Science from the University of Aberdeen. Before joining UNIMAS, he was an Operation Support Specialist at Hewlett Packard. He is best known for his work on sentiment analysis and opinion mining, automatically labelling sentiment-bearing topics, abstractive and extractive labelling, and opinion summarisation. Over the last few years, he has been interested in statistical techniques for natural language generation. His research in this area has included work in the subareas of text summarisation, topic modelling, and, more recently, opinion summarisation through word statistics, linear approach, and lexical resource acquisition through statistical means.

Second Author Noor Fazilla Abd Yusof is a senior lecturer in the department of Intelligent Computing & Analytics, Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM). She completed her PhD in Computing Science at the University of Aberdeen, Scotland. She has expertise in areas such as data analyt- ics, machine learning, artificial intelligence, and other related fields. Her research interest is in the area of intelligent computing and analytics such as machine learning, data analysis and analytics, and exploring the applications of AI in mental healthcare.

Third Author Donald Stephen is a senior lecturer in the i-CATS University College. He completed his PhD in Cognitive Science at the Universiti Malaysia Sarawak (UNIMAS). He has expertise in areas such as data analytic Psychometrics, Instrument Development, Behavioural Data Mining, Applied Statistics and Measurement. His research interest is in the area of Organizational Research, Community Research and Usability Studies.

Forth Author Mohd Kamal Othman is an Associate Professor in the Faculty of Cognitive Sciences and Human Development at the Universiti Malaysia Sarawak. He received his PhD in Computer Science (Human-Computer Interaction) from the University of York, UK and a Master’s degree in Virtual Environment: Technology and Practices from the University of East Anglia (UEA), UK. His research interest is multi-disciplinary, ranging from Cognitive Science, Human-Computer Interaction, Computer Science, Virtual Reality, Artificial Intelligence, and Learning Science and Education.

Fifth Author Zaimuariffudin Shukri Nordin has an educational background at the Faculty of Cognitive Sciences and Human Development. His research interests span the fields of education, sustainable development, and human rights. His current work examines the Circular Economy and links it to present practices while also investigating the Sustainable Development Goals at the Sarawak Region's Parliament.

Sixth Author Mukhriz Izraf Azman Aziz is an Associate Professor in the UUM College of Business at the Universiti Utara Malaysia. His research interests span the field of international economics, financial economics and econometrics.