MATIMYÁS MATEMATIKA
ISSN 0115-6926

Journal of the Mathematical Society of the Philippines
Vol. 45 No. 2 (2022) pp. 25-48

Time-Area-Household-Case Network Modelling of the COVID-19 Cases in Baguio City

Rosel R. Oryan
Department of Mathematics and Computer Science
University of the Philippines Baguio
Baguio City, Benguet
rroryan@alum.up.edu.ph

Joel M. Addawe
Department of Mathematics and Computer Science
University of the Philippines Baguio
Baguio City, Benguet
rcaddawe@up.edu.ph

Rizavel C. Addawe
Department of Mathematics and Computer Science
University of the Philippines Baguio
Baguio City, Benguet
rcaddawe@up.edu.ph

Jhunas Paul T. Viernes
Department of Mathematics and Computer Science
University of the Philippines Baguio
Baguio City, Benguet
jtviernes@up.edu.ph

Donnabel T. Panes
Baguio City Health Services Office
Baguio City, Benguet
dtpanes@slu.edu.ph

Abstract

In this study, a network is used to analyse the spread of COVID-19 in Baguio City from March 2020 to February 2021. A time-area-household-case model is derived from the dataset by generating nodes using four characteristics, namely, the *Onset of Illness* of a patient, residential *barangay*, household, and patient codes (*pcodes*) representing each COVID-19 case. The model is then used to establish a unimodal network to correlate case nodes and is measured using *degree*, *local clustering*, and *betweenness* among closely related nodes. The model has isolated certain cases with relatively interesting values from their neighbourhood, among which are cases in the beginning and later months of infection during the COVID-19 timeline in Baguio City.
Keywords: Network Modelling, Graph Theory, COVID-19
2020 MSC: 03C30, 05C12, 05C85, 05C90

1 Introduction

Coronavirus disease 2019 or COVID-19 is an infectious disease caused by the novel coronavirus SARS-cov-2, which is suspected of having originated from bats and pangolins and has subsequently been transmitted to humans through an intermediary host [1, 2]. The virus is present in the bodily secretions of the infected, which enables virus transfer through skin contact and transmission via sneezing, coughing, and talking. With water droplets as its carrier, the virus may linger in the air of an unventilated room or remain in a closed environment and on various surfaces for a period of time [3]. Common COVID-19 symptoms include coughs, shortness of breath, headache, fever, fatigue, and loss of taste or smell which may develop into more severe pneumonia, dyspnea, and hypoxia. It usually takes around 3 to 7 days for a person to show symptoms after contact with the infected. Some take up to two weeks or more [4, 5].

A COVID-19 patient is most infectious during the first week of showing symptoms. Studies suggest that patients become contagious around 2 to 4 days before manifesting the first signs of the disease, resulting in undetected presymptomatic infections. While asymptomatic patients cause silent transmissions, experiments have shown their influence to be minimal to none [6, 7].

With the help of current technologies and surveillance, it is now possible to track the spread of the virus in communities, trace contacts, and isolate infections to prevent further proliferation of the disease [8]. Among the different methods of investigating the spread of the disease, social networks have been used to analyse and visualise the growth of infections and interactions between social groups [9, 10].

A *network* is a set of nodes or points that are connected by a set of lines called edges. Networks model correlations in which its nodes correspond to entities, and its edges represent the relationship between nodes. Subsequently, multi-modal networks consist of two or more distinct sets of nodes. With the flexibility of its use, networks have been applied in criminal analysis, disease modelling, and structuring of real-world relationships [11, 12].

In this study, we attempt to use four factors: time, area, household, and individuals, to infer possible correlations between entities from a given set of finite data through the construction of a network model. In this case, we use the local COVID data to investigate cases of infection in the local community of Baguio City, Philippines, from March 2020, when the first infection in the city was detected, until February 2021.

2 Methodology

An epidemic occurs when an infectious disease has widely affected a population or region during a certain period. Given this definition, we note two factors when inspecting the spread of disease: people and their environment. Environmental factors include time and location. Social factors include individuals and their interacting communities or social groups.

Likewise, we now define the entities considered in our model — time, area, social groups and individuals. Suppose we consider the entities (time, area, household, individuals) as independent sets of interacting nodes, we map the relationship among nodes of the same set in the form of an adjacency matrix.

For instance, given a set of nodes N, its corresponding adjacency matrix is represented by $N_{adj} = [n_{b,d}]$ where $1 \leq b, d \leq |N|$ such that $n_{b,d} = 0$ if $b = d$; $n_{b,d} = \infty$ if no relationship exists between the nodes n_b and n_d; and $n_{b,d} = 1$ if a relationship exists between the nodes n_b and n_d, where $n_b, n_d \in N$. Hence, given the set of time nodes $T = \{t_1, t_2, \ldots, t_u\}$, area

nodes $A = \{a_1, a_2, \ldots, a_v\}$, household nodes $H = \{h_1, h_2, \ldots, h_w\}$, and nodes representing individuals $C = \{c_1, c_2, \ldots, c_x\}$, the corresponding adjacency matrices of each set is defined in Table 1.

Table 1: Summary of the General Conceptualisation of the Adjacency Matrices

$T_{adj} = [t_{i,j}] = \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,u} \\ t_{2,1} & t_{1,2} & \cdots & t_{2,u} \\ \vdots & \vdots & \ddots & \vdots \\ t_{u,1} & t_{u,2} & \cdots & t_{u,u} \end{bmatrix}$,	where $u =$ total no. of days being considered and $t_i, t_j \in T$ such that $1 \leq a, b \leq u$, $t_{i,j} = 1$ if $\lvert j - i \rvert = 1$, $t_{i,j} = 0$ if $i = j$, else $t_{i,j} = \infty$
$A_{adj} = [a_{k,l}] = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,v} \\ a_{2,1} & a_{1,2} & \cdots & a_{2,v} \\ \vdots & \vdots & \ddots & \vdots \\ a_{v,1} & a_{v,2} & \cdots & a_{v,v} \end{bmatrix}$,	where $v =$ total no. of distinct locations being considered and $a_k, a_l \in A$ such that $1 \leq k, l \leq v$, $a_{k,l} = 1$ if the $k^{th}$and l^{th} locations border one another (or if a direct footpath $\exists$ between them), $a_{k,l} = 0$ if $k = l$, else $a_{k,l} = \infty$
$H_{adj} = [h_{p,q}] = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,w} \\ h_{2,1} & h_{1,2} & \cdots & h_{2,w} \\ \vdots & \vdots & \ddots & \vdots \\ h_{w,1} & h_{w,2} & \cdots & h_{w,w} \end{bmatrix}$,	where $w =$ total no. of households under consideration and $h_p, h_q \in H$ such that $1 \leq q, r \leq w$, $h_{p,q} = 1$ if the $p^{th}$and q^{th} households are direct neighbours (adjacent), $h_{p,q} = 0$ if $p = q$, else $h_{p,q} = \infty$;
$C_{adj} = [c_{y,z}] = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,x} \\ c_{2,1} & c_{1,2} & \cdots & c_{2,x} \\ \vdots & \vdots & \ddots & \vdots \\ c_{x,1} & c_{x,2} & \cdots & c_{x,x} \end{bmatrix}$,	where $x =$ total no. of individuals under consideration and $c_y, c_z \in C$ such that $1 \leq y, z \leq x$, $c_{y,z} = 1$ if the $y^{th}$and z^{th} individuals are correlated (i.e. have had direct contact with one another), $c_{y,z} = 0$ if $y = z$, else $c_{y,z} = \infty$

Suppose we consider points of time and area, households, and individuals as interacting

sets of nodes, the relationship among nodes of different sets will be based on what causes these nodes to be correlated. In this case, infection.

Then suppose we define a *contagion* α as an event in which infections have occured. Like any event, every contagion would have a time of occurence and a location where the infection has taken place. We use $t_\alpha \in T$ to represent a day when α has occured and use $a(t_\alpha) \in A$ to represent the place affected by the contagion α during t_α. Likewise, households located in $a(t_\alpha)$ that have been affected by the contagion α during t_α may be represented by $h(a(t_\alpha)) \in H$. Lastly, patients affected by α may be represented as $c(h(a(t_\alpha))) \in C$. The purpose of these representations is to show the conditional relationship between nodes (Table 2).

Table 2: Illustration of Infections through Node Relationships

Node	Representation	Description
Time	t_α	Identifies a time when infections have taken place
Area	$a(t_\alpha)$	Identifies an area affected by infections during a particular period of time
Household	$h(a(t_\alpha))$	Identifies an infected household in an area during a given period of time
Case	$c(h(a(t_\alpha)))$	Specifies an infected individual

Patient records come with a variety of information, some of which may be isolated and used as data to map out infections. For example, a single data entry contains a patient's name, testing date, personal information, residence, etc.

To investigate COVID-19-related infections, we isolate three basic information from each entry: *Onset of Illness*, *Home Address*, and *Name*.

Onset of Illness is the date of when symptoms first manifested in symptomatic patients, or the date of when asymptomatic patients were tested.

Home Address indicates a patient's residential address. In this case, we only take the *barangay*[1] from each address.

Name identifies a patient. A patient's surname will serve as an identifier for the household in which they belong. The patient themselves are each assigned unique patient codes (*pcodes*) as case identifiers.

Note that all information used during the course of the study is masked under unique codes. For example, the first day under consideration (02-03-2020) is coded as t1, the area first affected by infections is coded as a1, the surname representing the first infected household is coded as h1, and the first recorded COVID-19 patient in the city is coded as c1. Codes are used for ease of data processing, for identifying the type of data (i.e. time or area) and its sequence in the overall timeline. Actual dates and names of persons and places are not visible. Dates, however, can be reverted given the visibility of the exact beginning of the timeline.

[1] *Barangay* is a territorial unit in the Philippines. Baguio City can be divided into territorial units called 'districts' and further into *barangays*.

Table 3: Example of a Single Data Entry

pcode	*Surname*	*Residence*	*Onset of Illness*
001	Dusslig	Military Cut-off	2 Mar 2020

Table 4: Example of a Single Data Entry Coded

pcode	*Surname*	*Residence*	*Onset of Illness*
c1	h1	a1	t1

2.1 Time-Area-Household-Case Model M

Based on the data structure used, we find a set of four entities $\{T, A, H, C\}$, with each subset defined in Table 5.

Table 5: General Representations of Node Entities

Set	Description
$T = \{t_1, t_2, \ldots, t_u\}$	Set of entities representing the u number of days t from 2 March 2020 to 28 February 2021
$A = \{a_1, a_2, \ldots, a_v\}$	Set of entities representing the v number of *barangays* a in Baguio City which have recorded COVID-19 cases throughout the considered timeline
$H = \{h_1, h_2, \ldots h_w\}$	Set of entities representing the w number of households h in *barangays* a which have been infected within the considered timeline
$C = \{c_1, c_2, \ldots c_x\}$	Set of entities representing the x number of COVID-19 patients c reported in the City

Area is based upon the residential address of each infected patient. While we cannot determine whether patients reside in the same residence, we can infer by matching their data. We deduce that infected people that have been ill at around the same time t, who belong to the same barangay a, and share the same surname h, are people that belong to the

same household. On the basis of this three-point similarity, we assume that infected cases satisfying these conditions indeed share a home. While this assumption may not entirely be true, we use this as a linking factor for the purpose of exploring the model's application. Other information may be used (i.e. street, building, subdivision, etc.), but since these information are not readily available, we use last name as a household indicator. Hence, household entities h, represented by surnames, can only be mapped to one area a each.

From the patient data, we generate virtual nodes for the network model implementing each node as a function β of four distinct data properties: time (Date of the *Onset of Illness*), area (*residence*), household-identifier (last name), and case-identifier (patient code or *pcode*; reinfected patients are given new *pcodes* to differentiate cases). In this way, all nodes are generalised as $\beta(p_1, p_2, p_3, p_4)$ and are segregated based on the properties they have; p_1 for time, p_2 for area, p_3 for household, and p_4 for case.

Table 6: Nodes as a function of properties

As a function	Description
$\beta(t, null, null, null)$	A day from the period 2 March 2020 to 28 February 2021
$\beta(t, a, null, null)$	*Barangays* affected by COVID-19 in Baguio City during a given time t
$\beta(t, a, h, null)$	A household under a particular area a that has been infected during t
$\beta(t, a, h, c)$	Infected individuals

We now define our network model as $M = \{N, E\}$ with a set of nodes N and edges E, such that $N \ni n$, where $n = \beta(p_1, p_2, p_3, p_4)$ and $p_1 \in T$, $p_2 \in A$, $p_3 \in H$, $p_4 \in C$; T, A, H, and C being a set of nodes gathered from the same dataset but are used as properties to generate the actual nodes used for the network model.

We also define an edge $e_{n,m} \in E$ as an edge connecting two nodes n and m where $n, m \in N$, $n \neq m$, $n = \beta(p_1, p_2, p_3, p_4)$ and $m = \beta(q_1, q_2, q_3, q_4)$ such that

$$e_{n,m} = \sum_{i=1}^{4} \begin{cases} Adj(p_i, q_i) & \text{if } p_i \neq q_i \text{ for } p_i, q_i \neq null \\ 1 & \text{if } p_i \neq q_i \text{ for } p_i \vee q_i = null \\ 0 & \text{if } p_i, q_i = null \end{cases}, \tag{1}$$

where $Adj(p_i, q_i)$ is the value of the adjacency between nodes p_i and q_i, from their predetermined adjacency matrix (as indicated in Table 1).

Given that the only known relationships are those that exist between time nodes and between area nodes, T_{Adj}, A_{Adj} are the only established adjacency matrices. H_{Adj} and C_{Adj} will only have values of ∞ or 0 if $p_i = q_i$ where $p_i, q_i \in H_{Adj}$ for $i = 3$ or $p_i, q_i \in C_{Adj}$ for $i = 4$.

As such, the edges may then be represented by an adjacency matrix N_{adj} mapping all nodes in N, such that

$$N_{adj} = [e_{i,j}] = \begin{bmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,s} \\ e_{2,1} & e_{1,2} & \cdots & e_{2,s} \\ \vdots & \vdots & \ddots & \vdots \\ e_{s,1} & e_{s,2} & \cdots & e_{s,s} \end{bmatrix},$$

where $1 \leq i, j \leq s$, $s = |N|$, and $e_{i,j}$ represents the value of the edge linking the i^{th} node to the j^{th} node in the network and is defined by Equation 1.

2.2 Small Scale Example

To illustrate, we model a network for data entries of four patients c_1, c_2, c_3, c_4 with data enumerated in Table 7.

Table 7: Sample Patient Data

pcode	*Surname*	*Residence*	*Onset of Illness*
c1	h1	a1	t1
c2	h2	a1	t1
c3	h3	a2	t2
c4	h3	a2	t2

From the sample data presented in Table 7, acquire a set of nodes $\{T, A, H, C\}$ each with the following elements:

$T = \{t1, t2\}$	time nodes
$A = \{a1, a2\}$	area nodes
$H = \{h1, h2, h3\}$	household nodes
$C = \{c1, c2, c3, c4\}$	case nodes

We model our network $M_{small} = \{N, E\}$ where $N \ni n$ such that $n = \beta(p_1, p_2, p_3, p_4)$ and $p_1 \in T$, $p_2 \in A$, $p_3 \in H$, $p_4 \in C$. N will be a set containing the nodes in Table 8.

Table 8: Nodes of the Small Scale Model M

Node Labels	As a function of Data Properties
n_1	$\beta(t1, a1, h1, c1)$
n_2	$\beta(t1, a1, h1, null)$
n_3	$\beta(t1, a1, null, null)$
n_4	$\beta(t1, null, null, null)$
n_5	$\beta(t1, a1, h2, c2)$
n_6	$\beta(t1, a1, h2, null)$
n_7	$\beta(t2, a2, h3, c3)$
n_8	$\beta(t2, a2, h3, null)$
n_9	$\beta(t2, a2, null, null)$
n_{10}	$\beta(t2, null, null, null)$
n_{11}	$\beta(t2, a2, h3, c4)$

Given the dataset, suppose T is a set of consecutive days correlated by the adjacency matrix T_{Adj}, where

$$T_{adj} = \begin{matrix} \\ \text{t1} \\ \text{t2} \end{matrix} \begin{matrix} \begin{matrix} \text{t1} & \text{t2} \end{matrix} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{matrix},$$

while the relationship between nodes in A, H, and C are unknown, therefore,

$$A_{adj} = \begin{matrix} \\ \text{a1} \\ \text{a2} \end{matrix} \begin{matrix} \begin{matrix} \text{a1} & \text{a2} \end{matrix} \\ \begin{bmatrix} 0 & \infty \\ \infty & 0 \end{bmatrix} \end{matrix}, \quad H_{adj} = \begin{matrix} \\ \text{h1} \\ \text{h2} \\ \text{h3} \end{matrix} \begin{matrix} \begin{matrix} \text{h1} & \text{h2} & \text{h3} \end{matrix} \\ \begin{bmatrix} 0 & \infty & \infty \\ \infty & 0 & \infty \\ \infty & \infty & 0 \end{bmatrix} \end{matrix}, \text{ and } \quad C_{adj} = \begin{matrix} \\ \text{c1} \\ \text{c2} \\ \text{c3} \\ \text{c4} \end{matrix} \begin{matrix} \begin{matrix} \text{c1} & \text{c2} & \text{c3} & \text{c4} \end{matrix} \\ \begin{bmatrix} 0 & \infty & \infty & \infty \\ \infty & 0 & \infty & \infty \\ \infty & \infty & 0 & \infty \\ \infty & \infty & \infty & 0 \end{bmatrix} \end{matrix}.$$

Each of the edges $e \in E$ would then be computed using Equation 1 to determine the values of the edges connecting all nodes in N, resulting to the adjacency matrix N_{Adj} as follows

$$N_{Adj} = \begin{array}{c} \begin{matrix} n_1 & n_2 & n_3 & n_4 & n_5 & n_6 & n_7 & n_8 & n_9 & n_{10} & n_{11} \end{matrix} \\ \begin{bmatrix} 0 & 1 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ 1 & 0 & 1 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & 1 & 0 & 1 & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 1 & 0 & 1 & \infty & \infty & \infty & \infty & 1 & \infty \\ \infty & \infty & \infty & 1 & 0 & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 1 & \infty & \infty & 0 & \infty & 1 & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & 0 & 1 & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & 1 & 1 & 0 & 1 & \infty & 1 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 1 & 0 & 1 & \infty \\ \infty & \infty & \infty & 1 & \infty & \infty & \infty & \infty & 1 & 0 & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & 1 & \infty & \infty & 0 \end{bmatrix} \end{array} \begin{matrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \\ n_7 \\ n_8 \\ n_9 \\ n_{10} \\ n_{11} \end{matrix} .$$

N_{Adj} is the matrix form of the network M_{small} and is represented in Figure 1, which shows only edges with finite nonzero values.

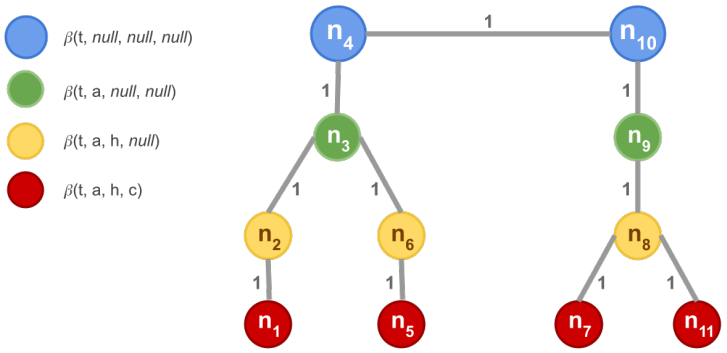


Figure 1: Small-scale Model

2.3 Modifying Time Properties

We consider the differences between asymptomatic, presymptomatic, and symptomatic patients relative to the collection of their samples for laboratory testing.

Asymptomatic patients are those that do not show COVID-19 symptoms.

Presymptomatic patients are infected patients that have not shown any COVID-19 symptoms at the time of testing but have developed symptoms after. They are recorded as asymptomatic.

Lastly, symptomatic patients are infected patients that have already begun showing signs of infection before testing.

Asymptomatic, presymptomatic, and symptomatic COVID-19 patients (those who do not independently avail for testing despite showing symptoms) are discovered through expanded testing (facility-wide testing of suspected groups of people or people highly at risk), through contact tracing (testing the contacts of a COVID-19-positive person) or miscellaneous testing (which are independent tests for travel purposes, medical operations, work

requirements, and other purposes). Only some symptomatic patients undergo voluntary testing as a result of the manifestation of symptoms.

Patients are considered infectious by the time they have begun manifesting symptoms. [13] However, for asymptomatic patients, this point in time remains vague.

Based on local data, *Onset of Illness* for symptomatic patients is the day when they first started showing their symptoms. For asymptomatic and presymptomatic, it is the day they are tested or one day before.

Other than the *Onset of Illness* the dataset used in the study also contains other time data such as the *Collection Date*, when testing was conducted, and *Release Date*, when results of the test were publicly announced. From the dataset, the number of days between each of these dates were computed for all patient entries; between the *Onset of Illness* and *Collection Date*, *Collection Date* to the *Release Date*, and from the *Onset of Illness* to the eventual *Release Date*. The computed general averages of each are shown in Table 9.

Table 9: Average No. of Days between Illness Onset, Sample Collection, and Result Release based on Local Laboratory Tests for Patients not Expired.

Conditions	Onset to Collection	Collection to Release	Onset to Release
Symptomatic patients excluding those tested due to symptoms	3.34	2.50	5.84
Symptomatic patients	3.37	2.36	5.73
Asymptomatic patients	0.04	2.11	2.16
Both symptomatic and asymptomatic patients	1.72	2.24	3.96

Discounting symptom-grounded tests, the mean length of time from the *Onset of Illness* of symptomatic patients to the collection of samples is 3.34 days.

Using the rounded delay in detection of symptomatic infections, we modify the *Onset of Illness* for asymptomatic patients to be 3 days before their testing date, based on the assumption that asymptomatic patients could not have suddenly developed the illness on the day they were tested. There are too many situations to account for in considering when a person may or might not have acquired then developed the illness. Hence, we use the average delay in detection for symptomatic patients who were involuntarily tested as a basis for standardising entries.

Additionally, based on the symptomatic averages, it takes around 6 days to detect a contagious patient who has not otherwise healed on their own. Suppose that right after a person is found to be infected, they are immediately quarantined and are monitored to prevent further spread of the disease. We assume a 1-day leeway from the release of results to the successful isolation of an infected individual. The resulting approximate 7 days from the *Onset of Illness* to isolation is a window of infection when contagious persons remain freely mobile throughout the city, not considering presymptomatic infectivity. This time range will be used in inferring the relationship of infections that could possibly be related.

2.4 Correlating Case Nodes

Earlier, the set of nodes T, A, H, and C were identified from the dataset in order to construct the network model M. We now investigate the otherwise unknown correlation between C nodes using the relationships established by M and aim to establish connections that do not yet exist between nodes of a particular node class. Specifically, we focus on case nodes to see how related a case is to the rest by creating direct links between nodes using pre-existing connections in the network M. The resulting projection would then be represented as a unimodal network $U = \{C, G\}$ where C is the set of case nodes and G is the weighted set of edges correlating each case node to another.

Note that the model $M = \{N, E\}$, and $N \ni n, n = \beta(p1, p2, p3, p4)$. If n has no properties that are *null*, that is, $p_1 \neq null$, $p_2 \neq null$, $p_3 \neq null$, and $p_4 \neq null$; and $p_4 = c$; then n will correspond to a particular node $c \in C \; \forall \; n$ that satisfies the previous conditions.

Hence given any two case nodes c_1 and c_2, where $c_1, c_2 \in C$, such that c_1 corresponds to a node n_1 and c_2 corresponds to a node n_2; we establish a correlation between c_1 and c_2 by creating an undirected edge $g \in G$ that connects n_1 to n_2 such that the weight of g will be the distance of the shortest path from n_1 to n_2.

2.4.1 Small Scale Example for Correlation

To illustrate, we use the small-scale network in Section 2.2 with nodes $N \in M_{small}$ listed in Table 8. In particular, the nodes $\{n_1, n_5, n_7, n_{11}\}$ are all generated from the function $\beta(p_1, p_2, p_3, p_4)$; p_1 to p_4 being properties derived from the *time*, *area*, *household*, and *case* segments of each data entry. The same segments are identified as nodes belonging to the sets T, A, H, C where $p_1 \in T$, $p_2 \in A$, $p_3 \in H$, and $p_4 \in C$, such that all nodes generated with non-*null* properties will each correspond to a node $c \in C$, where $c = p_4$. From the small-scale example, n_1 corresponds to $c1$, n_5 to $c2$, n_7 to $c3$, and n_{11} to $c4$, where the case nodes $C = \{c1, c2, c3, c4\}$.

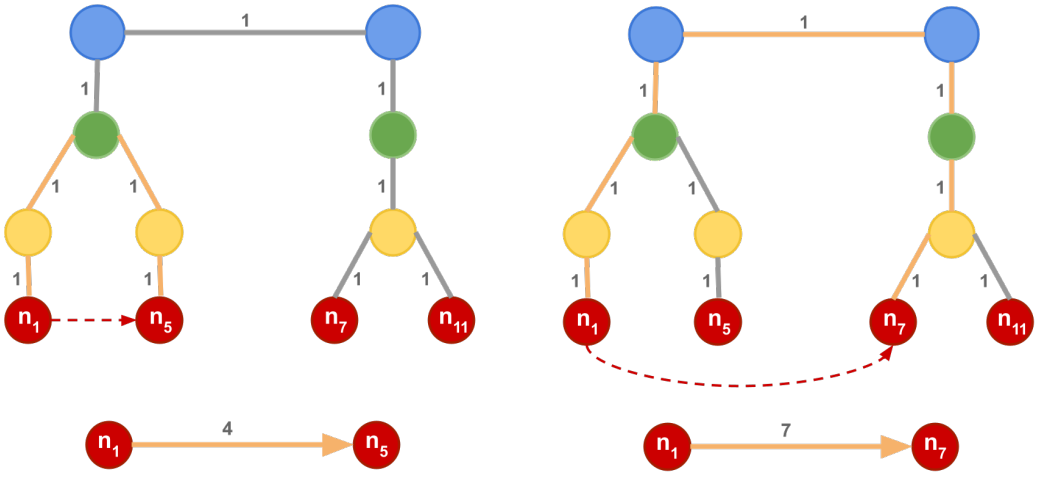


Figure 2: Small-Scale Correlation based on Example from Figure 1.

Shows the shortest-path links established from the first two correlations.

2.5 Limiting Correlations

Considering two cases c_1 and c_2, $c_1, c_2 \in C$ where c_2 was infected six months after c_1. Between the nodes that represent these cases, there would be a definite 180-day difference. Hence, we limit our projection such that c_1 and c_2 would be linked based upon a reasonable amount of time.

Symptoms can begin up to two weeks (14 days) after contact with an infected or sometimes, longer [14]. Hence we correlated case nodes $c \in C$ against other case nodes with time-points until 14 days after the 7-day contagion window.

In summary, we consider correlating a case node c_y to another case node c_z, where $c_y, c_z \in C$ and $1 \leq y, z \leq |C|$, only if the *Onset of Illness* of c_z is within 21 days after the *Onset of Illness* of node c_y.

2.6 Measuring the Network of Correlated Case Nodes

We use certain measures to rate the nodes in relation to their close neighbours rather than the entirety of the network. Given a unimodal network of case nodes $U = \{C, G\}$ where C is a set of case nodes and G is the set of edges correlating case nodes in C, a case node $c \in C$ will be measured for its *Degree*, *Local Clustering*, and *Betweenness*.

Degree will measure the number of direct neighbours c has, showing how many other people have been infected by COVID within 21 days before and after the *Onset of Illness* of c.

Node Degree

Node degree $deg(n)$ is the number of connections linked to a node n [12].

Node Clustering

Local Clustering $Clstr(n)$ measures to how interconnected n is to its neighbors, through the set of closed triplets $\tau_\triangle$ such that given any two nodes u and v both linked to n, the set of triplets $u - n - v$ would form a closed triangle if u and v are connected as well. Local Clustering is expressed as

$$clstr(n) = \frac{\tau_\triangle}{\tau},$$

where $\tau_\triangle$ is the number of closed triangles and τ is the total number of triplets in the neighbourhood of n [12].

Limited Betweenness

Betweenness $bet(n)$ is the measure of how much a node n serves as a bridge that connects some nodes u to v where $u \neq v$ via the shortest path such that

$$Bet(n) = \sum \frac{\sum d_{u \to n \to v}}{\sum d_{u \to v}} = \sum \frac{\text{sum of shortest paths from } u \text{ to } v, \text{ passing } n}{\text{sum of all shortest paths from } u \text{ to } v},$$

where $u \neq v \neq n$ for all nodes u to v in the network [12]. However, we apply betweenness measures to a subgraph around the vicinity of node n, determined by the distance d_{limit},

such that if the cumulative shortest path distance from any node to n exceeds d_{limit}, then we no longer consider it to be within the vicinity of n. We set this limit to $d_{limit} = 20$ (to make a comparison among data with at most a 20-point distance).

Local Clustering will measure how interconnected c is to its neighbours, while showing how clustered or how congregated infections are in a particular neighbourhood when comparing c to all other nodes within 21 days before and after its *Onset of Illness*.

Lastly, *Betweenness* will measure how much c serves as a connector to groups of infected cases within 21 days before and after its *Onset of Illness*. For example, given two sets of case nodes C_1 and C_2 where $C_1 \ni c \in C_2$, if c is a node with the closest and most numerous connections to nodes in both sets, and the resulting nodes from $C_1 \cap C_2$ is few, then c will have a high *betweenness*.

3 Results and Discussion

The first patient found to be COVID-19 positive was an overseas worker from Italy. The infected arrived in Baguio City on March 2, 2020. Hence time nodes begin on this day until February 28, 2021. During this span of time, there have been a total of 5560 COVID-19 Cases in the city.

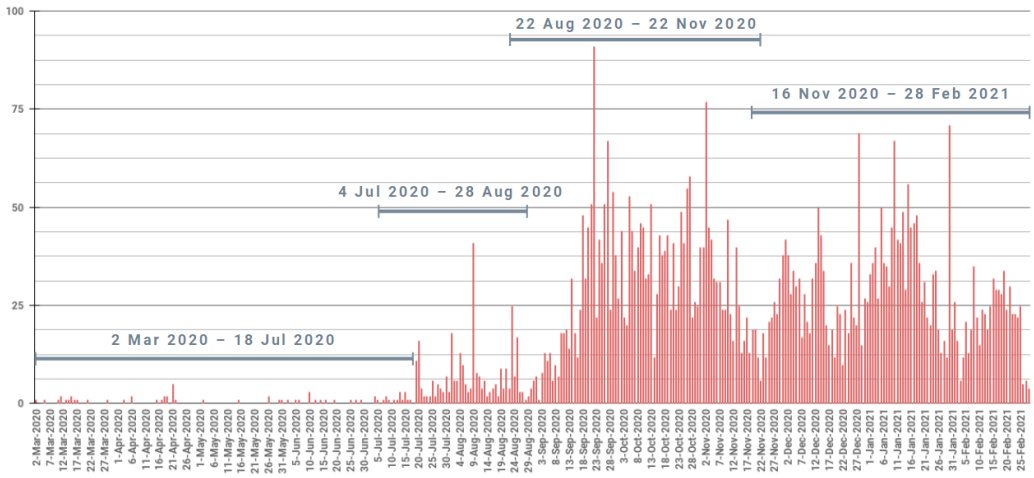


Figure 3: Daily COVID cases in Baguio City from 2 March 2020 to 28 February 2021 as generated from the dataset used in the study

Infections in the one year timeline of the COVID-19 epidemic (Figure 3) in the city are sectioned (as described in Table 10) based on the growth and decline of the disease. We generate a network for each section and analyse them one by one.

3.1 Comparison on Case Time Changes

Using the data from Phase 2 of the timeline, we compare the changes in the model and the resulting projection when the *Onset of Illness* of asymptomatic cases is pushed back by

Table 10: Phase Division in the One-Year COVID-19 Timeline

Timeline	Dates	Cases	Avg. case-per-day	Description
Phase 1	2-Mar to 18-Jul	65	1	Initial infections
Phase 2	4-Jul to 8-Aug	314	7	Moderate increase
Phase 3	22-Aug to 22-Nov	2650	30	Surge in cases
Phase 4	16-Nov to 28-Feb	2860	28	Maintained high freq.

three days.

The premise of shifting the nodes is to relate asymptomatic cases with uncertain *Onset of Illness* to previously occurring symptomatic cases with definite onsets, given that the *Onset of Illness* of symptomatic cases have always occurred before the testing date.

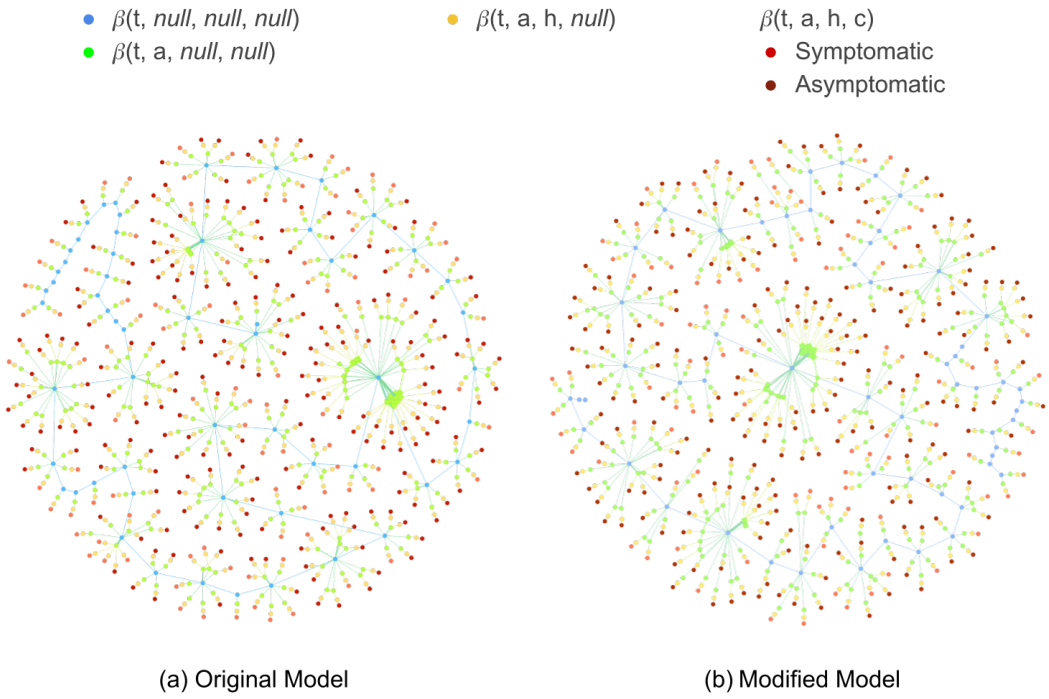


Figure 4: TAHC Model Phase 2 Comparison

When case nodes are shifted three days backward, the nodes, as well as their corresponding household and area nodes, are mapped to a different set of time nodes which, in turn, establishes different sets of links among adjacent area nodes and equal household nodes. The difference is not evident in the visualisation of TAHC models (Figure 4) in the presence of the multitude of data points. Hence, we evaluate the effects of the onset modification reflected in their corresponding projections.

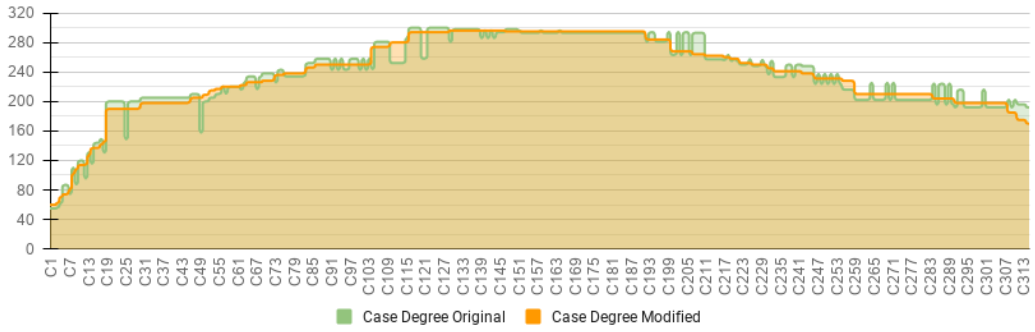


Figure 5: Case Node Degree of Original and Modified

Phase 2 Models for correlated Case Nodes

Phase 2 begins on July four which we locally set as Day 1. This particular phase of the dataset contains 314 cases, 218 of which are asymptomatic cases while the lesser remaining are symptomatic. Therefore, the shift in node positioning involves the majority of nodes in the network.

The highest degree value in the original model is 300, determined by nodes mapped to day 35 (all asymptomatic), followed by a degree of 296 determined by nodes mapped to day 36 (asymptomatic and symptomatic). In the modified model, the highest degree value is 396, determined by nodes mapped to days 33, 34, and 35, followed by a degree of 295 determined by nodes mapped to days 36 and 37 (both asymptomatic and symptomatic as shown in Figure 5).

None of the patients tested have had their symptoms begin on day 35 and after the onsets shifted, the nodes once mapped to day 35 moved to day 32. Since these nodes hold the most degrees, when they are moved, they carry their connections along.

Meanwhile, the lowest degrees are 55, 56, and 62 from the original model and 60, 62, and 71 from the modified model, similarly determined by nodes mapped to days 1, 2, and 4. The cases represented by these nodes are symptomatic cases with fixed onset dates with degrees influenced by the shifted nodes from subsequent days.

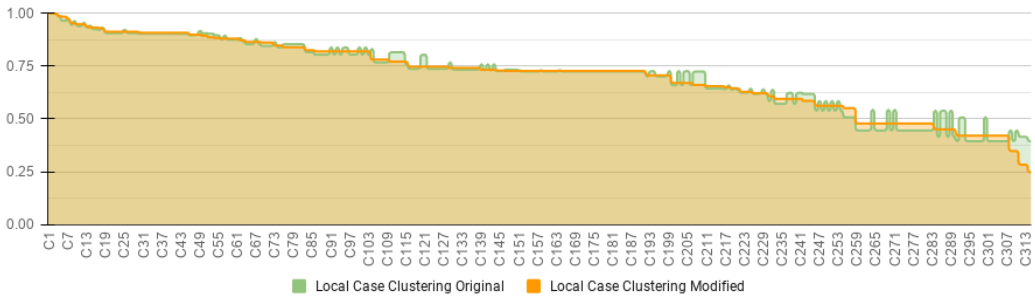


Figure 6: Local Case Node Clustering of Original and Modified

Phase 2 Models for correlated Case Nodes

Both models show greater clustering values (0.9 to 1) among nodes at the beginning of

the Phase 2 timeline and subsequent lower local clustering values (0.2 to 0.5) for nodes at the end of the timeline (Figure 6).

On the other hand, the Phase 2 models with the highest betweenness values are case nodes mapped to day 40 while in the modified model, the highest betweenness values are case nodes mapped to day 37, of which are nodes within the range of C150 to C210 for both models (Figure 7).

The influence of the 3-day-pushback in *Onset of Illness* upon the connections in the network for correlated case nodes is proportional to the amount of asymptomatic nodes moved at a particular time point. The overall connections among asymptomatic nodes are maintained while their relative links to prior symptomatic cases are strengthened and links to succeeding symptomatic nodes are loosened.

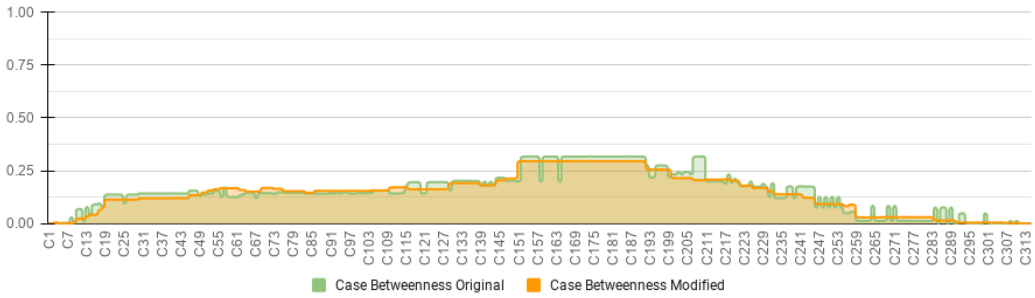


Figure 7: Case Node Betweenness of Original and Modified Phase 2 Models for correlated Case Nodes

3.2 Correlating Case Nodes

The dataset is modified such that the *Onset of Illness* is pushed back 3-days from the testing date on the basis of the average time it takes a symptomatic patient to be tested after the onset of their symptoms. Using the model constructed from this dataset, we derive a relationship among case nodes by projecting the model in which we examine certain case nodes that reflect interesting values.

There are a total of 5560 cases, each represented by case nodes $C = \{C1, C2, \ldots, C5560\}$.

3.2.1 Phase 1

From March 2 to July 18 of 2020, cases C45, C46, and C47 have shown the greatest degree ($deg = 28, 27$) followed by C48, C49, and C50 ($deg = 23$), while the case with the least degree is C30 ($deg = 7$), as shown in Figure 8 (a).

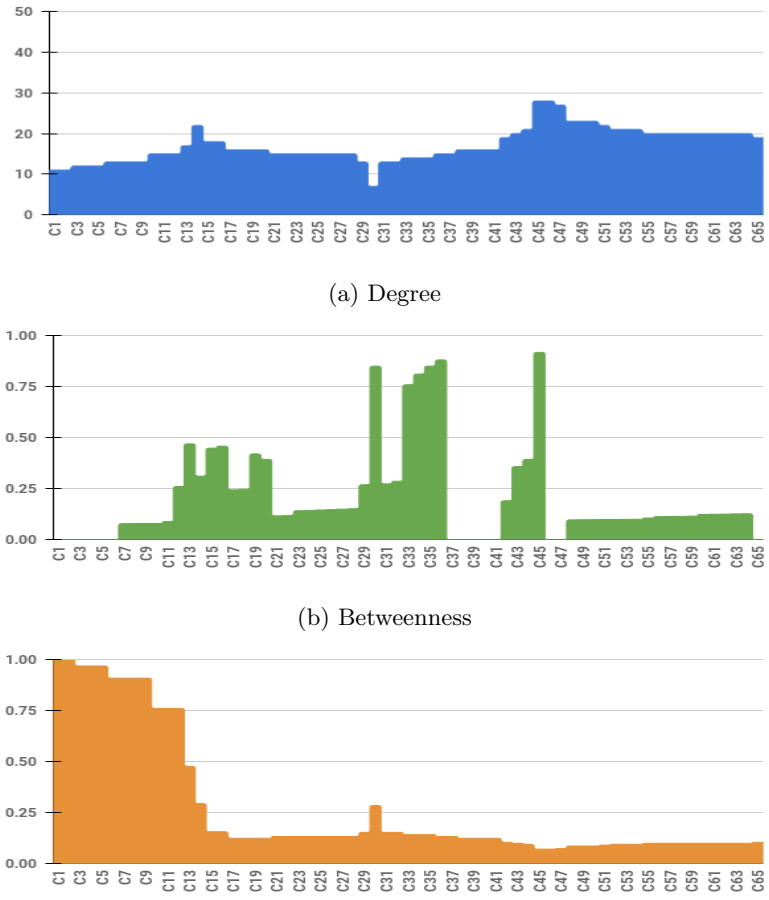
Among betweenness values, shown in Figure 8 (b), the highest is from case C45 ($bet = 0.919$), followed by C30, C34, C35, and C36 ($0.8 \leq bet \leq 0.9$). The lowest betweenness of 0 is from cases C1 to C6, C37 to C41, C46, C47, and C65.

Regarding nodal clustering, shown in Figure 8 (c), cases C1 to C9 have shown the highest values ($0.9 \leq clstr \leq 1.0$) while C44 to C54 have the lowest ($0.7 \leq clstr \leq 0.9$).

In particular, case C45 has numerous links and its neighbourhood is highly interconnected. Its high betweenness implies locally relevant connections. We also have case C30 with few connections and a loosely linked neighbourhood but has a high betweenness value

which is likely due to C30 having been affected first prior to several infections that have occurred shortly after.

Lastly, we have cases C33 to C36 with degrees of around 14 to 15 (which are within the overall average degree count for Section 1, ranging from 14 to 17), low clustering values ($clstr \approx 0.14$), and high betweenness ($0.7 \leq bet \leq 0.9$). Cases C33 to C36 are numbered consecutively, implying that these cases have occurred successively or at around the same time.



(a) Degree

(b) Betweenness

(c) Clustering

Figure 8: Node Degree, Betweenness, and Clustering of the Phase 1 Model for correlated Case Nodes

Figure 9 shows the correlated case nodes in the unimodal case-node network for Phase 1, where darker shades indicate high betweenness values among case nodes and lighter shades indicate lower betweenness among nodes.

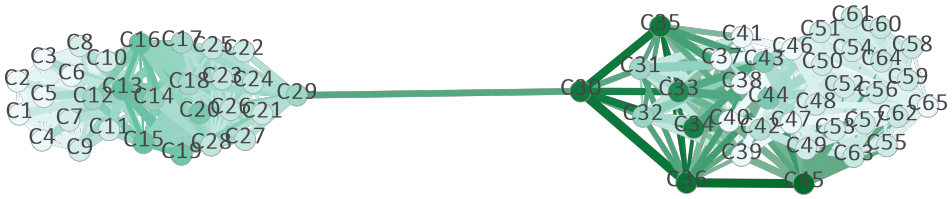
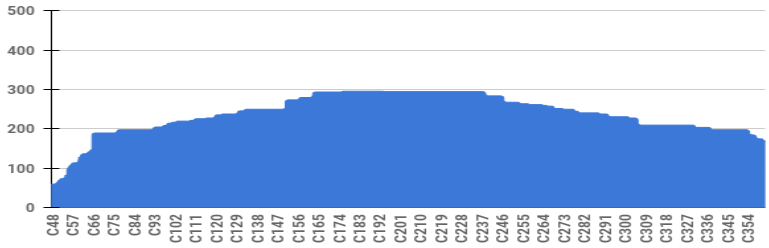
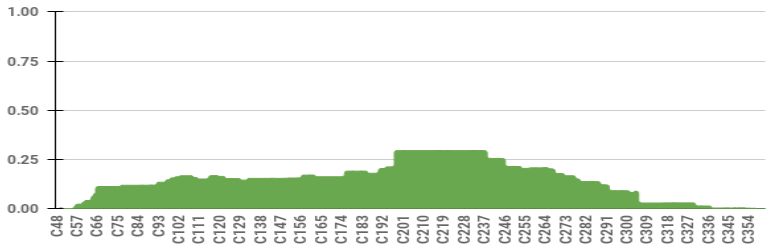


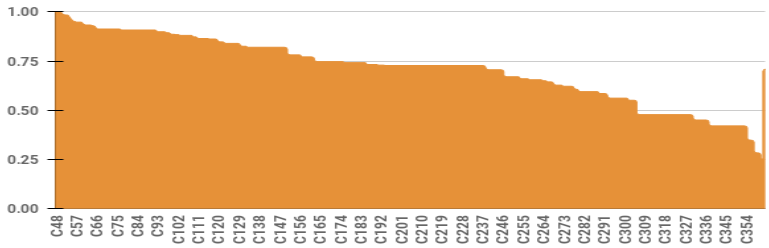
Figure 9: Phase 1 Model highlighting Limited Betweenness for correlated Case Nodes



(a) Degree



(b) Betweenness



(c) Clustering

Figure 10: Node Degree, Betweenness, and Clustering of the Phase 2 Model for correlated Case Nodes

3.2.2 Phase 2

From July 4 to August 28 of 2020, the network shows greater node degrees (shown in Figure 10 (a)) compared to Phase 1 of the COVID-19 timeline. From an average of 1 case per day in Phase 1 to 7 cases per day in Phase 2, the increase in cases contributes to the growth in nodal interconnections with degrees in the range of $7 \leq deg \leq 28$ in Phase 1 and $60 \leq deg \leq 296$ in Phase 2.

Similarly, the rise in clustering values of Phase 2 (average $clstr \approx 0.709$) in comparison to Phase 1 (average $clstr \approx 0.272$) indicates growth in node interconnection brought about by the increase in subsequent infections (refer to Figure 10 (b)).

On the other hand, betweenness values occupy a smaller and lower range in Phase 2 ($0 \leq bet \leq 0.30$) as compared to Phase 1 ($0 \leq bet \leq 0.92$) showing that the nodes in Phase 2 are more evenly positioned among closely related neighborhoods (refer to Figure 10 (c)).

3.2.3 Phase 3

From August 22 to November 22 of 2020, node degrees increase (average $deg \approx 1335$) with the surge in COVID-19 cases while the overall clustering values decrease throughout time; clustering being $0.9 \leq clstr \leq 1$ for the first 59 cases with degrees $202 \leq deg \leq 278$ for infections that have occurred within days 1 to 6 of Phase 3. Node neighbourhoods derived from the first six days show highly interrelated cases (Figure 11 (a) and (c)).

Clustering values decrease to $0 \leq clstr < 0.2$ for cases from day 46 onward for the last 1209 case nodes with degrees ranging at $616 \leq deg \leq 1665$. The neighbourhoods in this time range are less interconnected while having more numerous links.

Betweenness on the other hand ranges from $0 \leq bet < 0.33$, the highest values being $0.30 \leq bet < 0.33$ from cases C406 to C504 ($bet \approx 0.301$) occurring at day 22 of Phase 3, and cases C1581 to C1696 ($0.30 < bet < 0.33$) which are infections during days 51 to 53.

3.2.4 Phase 4

From November 16, 2020 to February 28, 2021 (refer to Figure 12), the overall node degrees decrease (average $deg \approx 1147$), falling within the range $519 \leq deg \leq 1445$ while the measured node betweenness range from $0 \leq bet < 0.35$.

Meanwhile, clustering values begin with clustering values $0.90 \leq clstr \leq 1$ for cases in the first 13 days of the timeline and declines to $0 < clstr < 0.2$ from day 43 onwards.

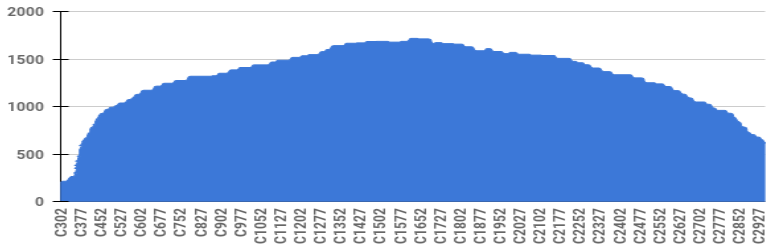
We highlight certain nodes holding values of interest such as class nodes with degree and betweenness measure values among the highest or lowest in Phase 4 of the timeline.

We define link importance to be the likelihood of links to have contributed to contact infections, which we infer preliminarily using betweenness values. Samples 1 and 2 from Table 11, represent nodes with the highest betweenness and degree values or cases with the most important links in the neighborhood of nodes mapped to days within or close to days 42 and 43 of Phase 4.

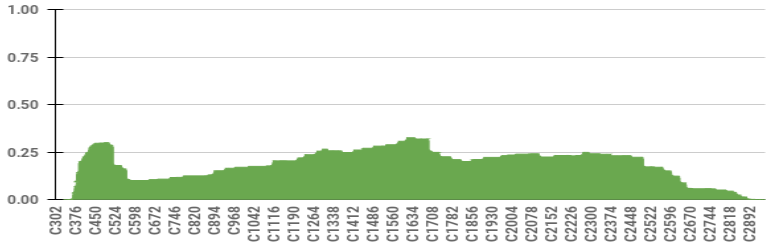
Samples 3 and 4 represent nodes with the highest degree but moderate betweenness overall. However, they hold the most important links among nodes mapped to days within or close to days 47 and 55.

Lastly, Sample 5 represents a section of nodes that have the least connections and the least significant links given their betweenness value of zero.

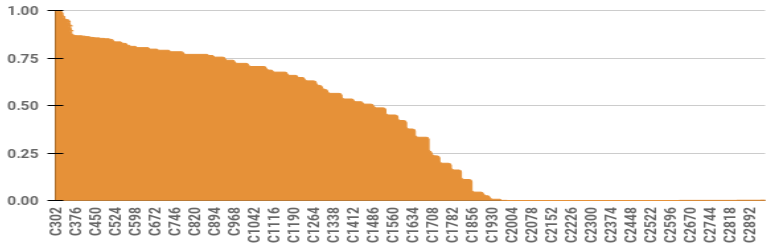
Interestingly, the cases represented by these nodes all belong to the same area (area code A7) and the same day (D56) which provides these nodes with short direct links among one



(a) Degree



(b) Betweenness



(c) Clustering

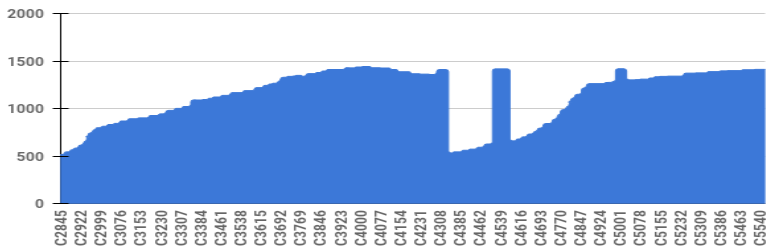
Figure 11: Node Degree, Betweenness, and Clustering of the Phase 3 Model for correlated Case Nodes

another. Zero betweenness value indicates that there exists a much shorter set of links that interconnect these nodes indirectly.

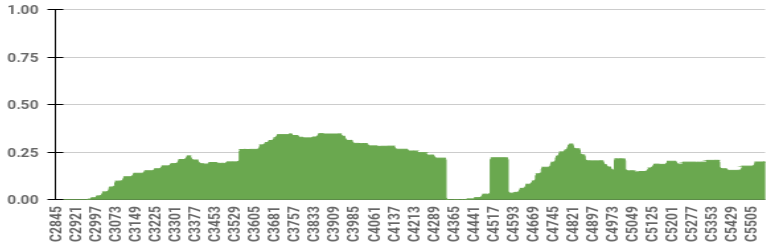
Figure 13 shows the correlated case nodes in the unimodal network for Phase 4, where darker shades indicate high betweenness values among case nodes and lighter shades indicate lower betweenness among nodes. Note, however, that nodes at the ends of the timeline are of a lighter shade because connections to nodes before and after the timeline of Phase 4 were not considered.

4 Conclusion

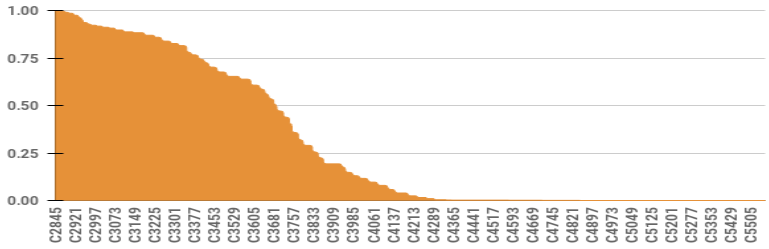
The weight between nodes that do not share similarities in their data are assigned values relative to how far apart infections are in terms of time. Hence, infections represented by



(a) Degree



(b) Betweenness



(c) Clustering

Figure 12: Node Degree, Betweenness, and Clustering of the Phase 4 Model for correlated Case Nodes

case nodes, such as C30 and C33-C36, which have taken place shortly before the occurrence of multiple infections reflect values significantly distinct from its neighbourhood. However, as infections become more frequent, the network becomes much more interconnected. As the network grows, nodes become more clustered and show more similar than unique interconnections, resulting to groups of nodes holding similarly high or low measures rather than each node being individually segregated, as displayed in Phase 3 and 4. On the other hand, Phase 1, which holds a smaller dataset, singles out distinct nodes rather than clusters.

The relationship modelled through the TAHC network relies mostly on time difference to create *distance* between nodes. Although the area and household similariy between data allows nodes to be grouped, time has the most influence on edge weight when establishing unimodal relationships.

The network model uses the time-area-household-case conditions to identify relationships from a pool of information, similar to a search criteria that specifies requirements of a

Table 11: Phase 4 Nodes of Interest

Label	Period	Case Nodes	Degree	Betweenness
Sample 1	Day 42	C3851 to C3870	1398	0.348
Sample 2	Day 43	C3871 to C3939	1414	0.347
Sample 3	Day 47	C4008 to C4040	1445	0.296
Sample 4	Day 55	C4295 to C4339	1414	0.219
Sample 5	Day 56	C4352 to C4357	535	0

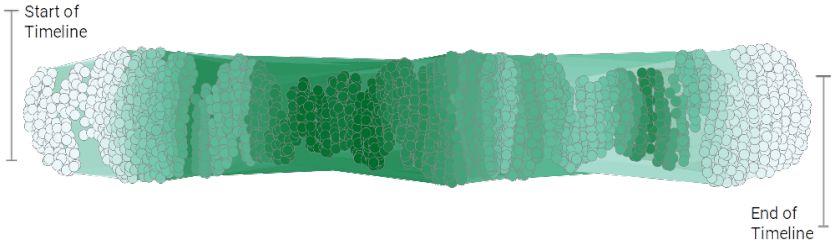


Figure 13: Phase 4 Model highlighting Limited Betweenness for correlated Case Nodes

product. Its tiered structure similar to a network tree allows the ranking of correlations between entities based on known similarities via a shared characteristic which we represent in the form of node properties.

Using other properties of an entry from the available dataset might lead to more successful grouping of relationships; the recommended criterias being workplace, shifts (i.e. active working hours), age (which may determine activity throughout the day) or other relating factors. Additionally, properties with more concretely determined relationships among nodes of the same kind are recommended. The model may also be extended to more than four properties to reduce vagueness or shorten the number of properties used to reduce specificity.

Overall, networks that aim to model real-world relationships, more often than not, result in complex and convoluted correlations. Likewise, the TAHC model. While more nodes and connections incorporated in the network makes for a more realistic model, it becomes just as difficult to interpret. But in this way, projection or conversion methods can select and isolate relevant relationships for analysis and serve as transformative measures that evaluate existing multimodal relationships.

Lastly, the conversion of its multi-layered relationships to a unimodal network simplifies such complexities into something more readable but will result in the framing of the model into only one perspective. While the model may be shown in a way that allows one to obtain a coherent series of relationships, other possibilities of interpretation will no longer be considered. That is, the original TAHC model can be used to establish correlations between area nodes or household nodes instead of case nodes by shedding away the properties that succeed it, but will lose this possibility once converted to a unimodal network consisting only of case nodes.

Acknowledgements

We would like to thank the Baguio City Health Services Office for allowing us use of the City's COVID-19 data.

References

[1] World Health Organization. Who-convened global study of origins of sars-cov-2: China part. 2021.

[2] Yen-Chin Liu, Rei-Lin Kuo, and Shin-Ru Shih. Covid-19: The first documented coronavirus pandemic in history. *Biomedical Journal*, 43(4):328–333, 2020.

[3] World Health Organization. Transmission of sars-cov-2: implications for infection prevention precautions. 7 2020.

[4] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application. *PubMed*, 05 2020.

[5] Zhanwei Du, Xiaoke Xu, Ye Wu, Lin Wang, Benjamin Cowling, and Lauren Ancel Meyers. Serial interval of covid-19 among publicly reported confirmed cases. *Emerging Infectious Disease journal*, 2020.

[6] Jennifer Bender, Michael Brandl, Michael Höhle, Udo Buchholz, and Nadine Zeitlmann. Analysis of asymptomatic and presymptomatic transmission in sars-cov-2 outbreak, germany, 2020. *Emerging Infectious Disease journal*, 2021.

[7] Andrew A Sayampanathan, Cheryl S Heng, Phua Hwee Pin, Junxiong Pang, Teoh Yee Leong, and Vernon J Lee. Infectivity of asymptomatic versus symptomatic covid-19. *The Lancet*, 2021.

[8] Sera Whitelaw, Mamas A Mamas, Eric Topol, and Harriette G C Van Spall. Applications of digital technology in covid-19 pandemic planning and response. *The Lancet Digital Health*, 2020.

[9] S Saraswathi, A Mukhopadhyay, H Shah, and T S Ranganath. Social network analysis of covid-19 transmission in karnataka, india. *PubMed*, 2020.

[10] Parul Maheshwari and Réka Albert. Network model and analysis of the spread of covid-19 with social distancing. *Applied Network Science*, 2020.

[11] International Association of Crime Analysts. Social network analysis for law enforcement. 2 2018.

[12] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[13] Centers for Disease Control and Prevention. Interim guidance on ending isolation and precautions for adults with covid-19. 3 2021.

[14] Marco Cascella1, Michael Rajnik, Abdul Aleem, Scott C. Dulebohn, and Raffaela Di Napoli. Features, evaluation, and treatment of coronavirus (covid-19). 4 2021.

This page is intentionally left blank