*Michael Gebreslasie*

**Track virus spread with ArcGIS Insights.**

In this session, we will use *link analysis* to trace contacts as a virus spreads through a community.

A community is experiencing an outbreak of an infectious disease. As a contact tracer for the local public health office, you must use the case data to determine where the disease spread, who requires testing, and which cases require further investigation. You will use ArcGIS Insights to create charts and apply nonspatial analytical techniques to track the spread of the virus.

*This tutorial is part of the spatial epidemiology module (underdevelopment).*

**Requirements**

* ArcGIS Insights
* ArcGIS organizational account with a Publisher, Facilitator, or Administrator role in an ArcGIS organization. *Note: see options for software access*: (<https://www.esri.com/en-us/arcgis/access-options?rsource=https%3A%2F%2Fwww.esri.com%2Fen-us%2Flg%2Ftraining-and-services%2Flearn-arcgis-education-trial>)
  + *Higher education with the university email address authorised by the university software administrator*.

**Outline (outcome of this practical)**

* Create charts for contract tracing: create charts that can be used to trace the spread of a virus through a community.
* Interpret relationships between cases: Use the charts to interpret how the virus has spread and make decisions about which cases require follow-up or testing.

**Create charts for contact tracing.**

You will load and analyse example data from a community that is experiencing a viral outbreak. Due to the sensitive nature of patient data, the data used in this analysis is based on a fictional community. The data was created with an assumption that each case infects approximately two new cases, unless measures are taken to isolate after infection. The data also assumes that male and female cases are equally likely to be infected. The age of infected cases is affected by the locations where the cases interacted, rather than certain age groups being more susceptible to infection. You will create charts that will allow you to perform contact tracing in the community.

**Open an ArcGIS Insights workbook.**

You will start by downloading a workbook package (*copy from USP and save it in your local drive*) and importing it into ArcGIS Insights.

1. Save the *community outbreak insightswbk* file in your local drive Download the.

The item is an Insights workbook package containing patient case data and locations where each case had close contact with other people.

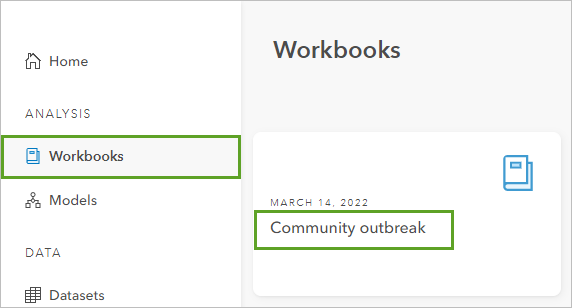
1. Go to the [Insights](https://insights.arcgis.com/#/) home page. Or sign in to the ArcGIS Online Portal - below: <https://www.arcgis.com/index.html>
2. Sign in to your [ArcGIS organizational account](https://www.arcgis.com/home/signin.html).
3. If necessary, click Skip in the Welcome to Insights window.

Next, you will import the package you downloaded.

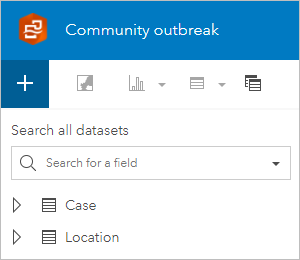
1. On the home page, click the Workbooks tab.
2. Click Import, browse to the workbook you downloaded, and double-click Community\_outbreak.insightswbk to add it to Insights.

The workbook is loaded into Insights and appears on the Workbooks page.

1. On the Workbooks page, click Community outbreak to open the workbook.



The workbook contains two datasets, *Case* and *Location*, in the data pane. Both datasets are stored in Excel tables and were added to the workbook that you imported. Insights supports data from many sources, such as Excel, feature layers from ArcGIS Online, or data from ArcGIS Living Atlas of the World.



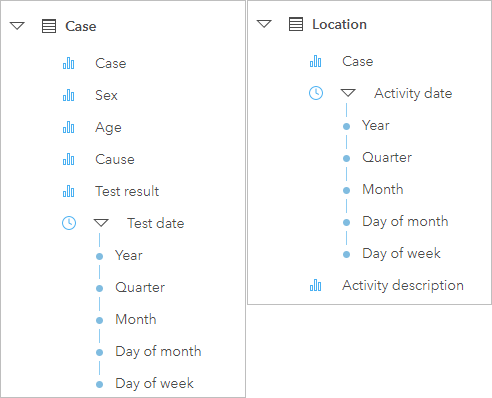
The workbook is titled Community outbreak and the page is named Contact tracing.

Next, you will use the datasets included in the workbook to create charts that display how the virus has spread through the community.

**Create a link chart to trace contacts.**

Excel tables, like the datasets saved in the workbook package, are not spatially enabled when added to Insights. There are ways to quickly enable location on your data; however, since this data is not based on a real community, you will focus this workflow on nonspatial charts. You will start with a link chart, which will show the connections between the cases in the community.

1. In the data pane, click the arrow next to both the Case dataset and the Location dataset to show the fields.



**Case** includes the fields Case, Sex, Age, Cause, Test result, and Test date. **Location** includes the fields Case, Activity date, and Activity description.

The Case field refers to people who are being tested, or who are being monitored because of close contact to someone who is being tested. The Case dataset lists each case once and provides details on each of the cases. Cases are identified by their household, sex (male, female, or other/prefer not to specify), and age range.

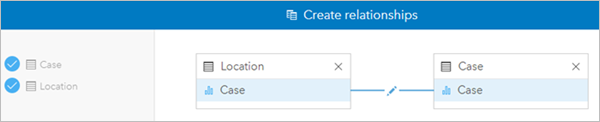
The Location dataset tracks the activities that each case participated in or locations where contacts were made in the days preceding testing.

To track viral infections, you will want to see the relationship between cases (including the cause of infection and test results) and reported contact locations. Since the information about each case is stored in a separate table from the reported activities, you will start by joining the two datasets. The joined dataset will be used to create most of the charts in your analysis and will allow all the charts to interact with each other, even when they reference data from different datasets.

1. Above the data pane, click Create relationships.

Create Relationship button

In the Create relationships window, click Location and Case.



The datasets are added to the window. Joins must be made based on shared attributes in the two datasets. For your data, the Case fields were chosen as the joining fields, meaning features with the same case value will be joined.

By default, Insights performs inner joins, which means that features that match for both datasets are included in the joined dataset. You want to keep the Location dataset as it is and add the case information to each location entry, so you will change the join type to a left join.

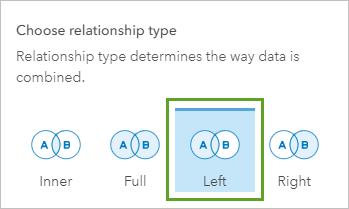
1. On the line joining the two datasets, click the edit button.



In the Edit relationship window, you can choose to keep all data from both datasets, all data from the left dataset, or all data from the right dataset. The Location dataset is on the left side of the join, so that’s the join type you will use.

*Note: If you added the Case dataset first, so that Location is on the right side of the join, you must choose Right as the join type instead.*

1. In the Edit relationship window, under Choose relationship type, click Left.



1. In the Create relationships window, click Finish.
2. In the data pane, click the arrows beside the datasets Case and Location to collapse them.

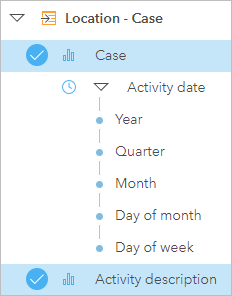
The joined dataset, called Location – Case, is added to the data pane. Joined datasets and other result datasets are indicated with an orange icon.



1. In the data pane, expand Location – Case.

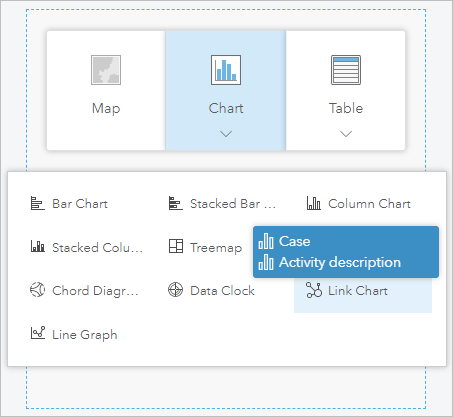
Now, you will use the fields in the joined dataset to make charts that you can use for contact tracing.

1. Click the circles next to Case and Activity description to select both fields.

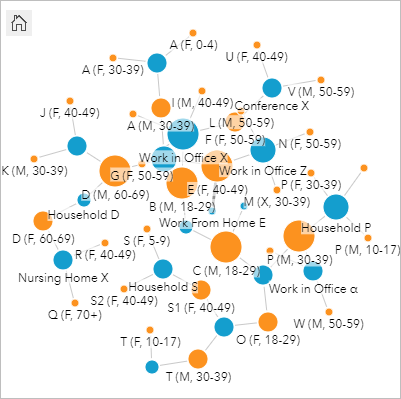


The two fields are selected in the data pane and ready to be added to the card type of your choice. You add fields to cards by dragging them onto the empty page to access the card options.

1. Drag the selected fields onto the page and point to Chart to see the options.



1. Drop the fields onto Link Chart.



A link chart is created showing the cases and locations, as well as the connections between them. You will change some settings on the link chart to optimize your analysis.

1. If necessary, click the link chart to activate it.
2. On the toolbar, click the Layer options button.

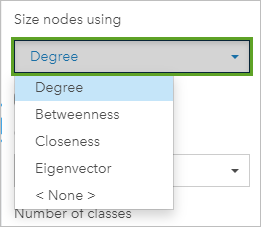
Legend button

The Layer options pane displays properties for the chart.

The nodes, which represent the cases and activity descriptions, are sized by graduated symbols using a method called centrality. The default centrality is measured by degree (the centrality measure appears in the Layer options pane, on the Symbology tab, under the Graph options parameter).

The degree refers to the number of links a node has. A person who reported many activities or a place that many people reported visiting will have a larger degree than a person who reported few activities or a place with a small number of people reported visiting.

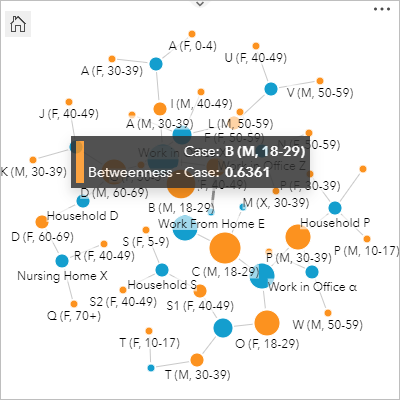
1. Under Size nodes using, expand the menu to see the other centrality options.



The following centrality options are available:

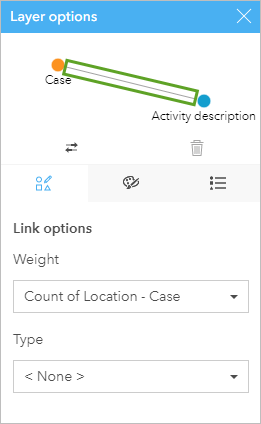
* + Degree—The number of direct neighbours of the node.
  + Betweenness—The extent to which a node lies on the shortest path between other nodes in the network.
  + Closeness—The average of the shortest distance paths to all other nodes.
  + Eigenvector—The measure of the influence of a node in a network based on its proximity to other important nodes.

1. Click Betweenness to change the centrality measurement.
2. Hover over a node on the chart.



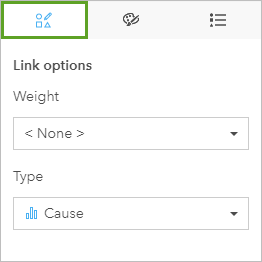
The hover pop-up tells you the case and betweenness for the node you pointed to. By default, the centrality is normalized so the values range from 0 to 1. Next, you will configure the links on the chart.

1. In the Layer options pane, select the link connecting Case to Activity description.



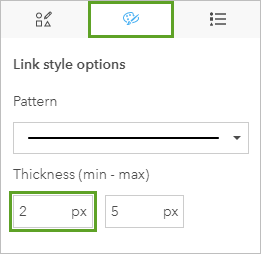
The Graph options are replaced with Link options.

1. For Weight, select <None>. For Type, select Cause.



The links are styled using unique values based on the Cause field. The lines are thin, so it can be difficult to see which colours are used. You will change the line thickness.

1. Click the Appearance tab. For Thickness (min - max), change the minimum value to 2 px.

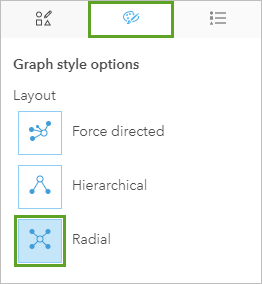


Finally, you will configure the layout of the link chart.

1. In the Layer options pane, click in an empty space near the link to deselect it.

The Graph style options appear in place of the Link style options.

1. For Layout, choose Radial.



The radial layout displays the nodes with the highest centrality near the centre, with links directed outward in an orbital pattern.

1. Close the Layer options pane. On the workbook toolbar, click Save.

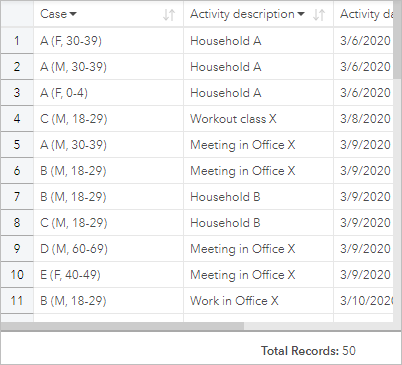
Save button

You have created a relationship between two separate datasets containing cases and locations and then created a link chart. Next, you will create charts and tables that display how the virus has spread through the community.

**Create charts to analyse cases.**

The link chart created in the previous step shows the relationship between cases using shared locations. In this step, you will create charts and tables that will give you more information about the cases and the activities they reported prior to their test.

1. In the data pane, expand ***Location - Case***.
2. Select Case, Activity description, Activity date, and Test result. Drag the fields to an empty area on the page, and drop them on Reference table.



A reference table is created showing the activities and test results for all cases.

The default card size is not wide enough to show all of the columns on the table. You will increase the card size.

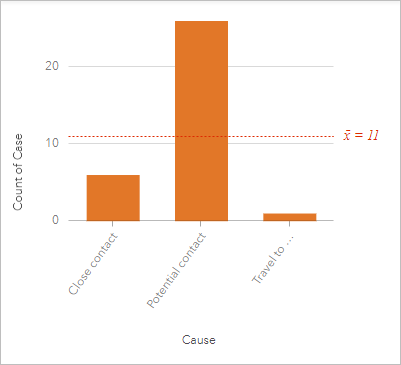
1. Use the side handles to increase the width of the card so all columns are visible.

The columns are all visible, but you won't be able to see all the rows at once. To make the reference table more usable, you'll apply a cross filter. The cross filter will filter the entries on the reference table based on selections made on other cards.

1. On the reference table toolbar, click the Enable cross filters button.

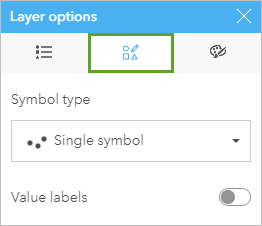
Next, you will create a column chart to show the cause of the potential infection for the cases. You will use the Case dataset rather than the joined dataset to ensure each case is only being displayed once.

1. In the data pane, expand Case.
2. Select Cause, drag it to an empty area on the page, point to Chart, then drop the field on Column Chart.

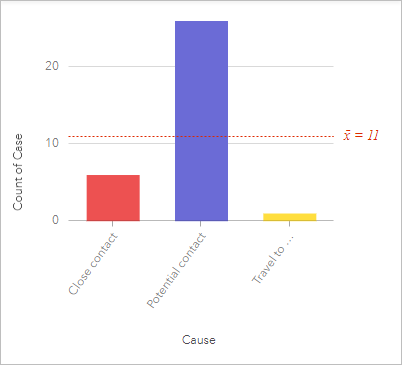


The column chart shows each cause (Close contact, Potential contact, and Travel to infected area) and the number of cases categorized for each. The columns on the chart are all the same colour. You will change the symbol to match the colours for the links on the link chart.

1. Click the column chart to activate it, then click the Layer options button to open the Layer options pane.
2. Click the Symbology tab.



1. For Symbol type, choose Unique symbols.

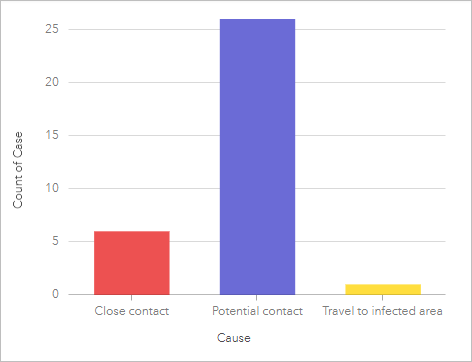


The chart is styled by unique symbols, using the same colours as the link chart. The chart also shows the average count. You do not need the average count, so you will remove it.

1. Click the Chart statistics button. In the Chart statistics pane, uncheck Mean.

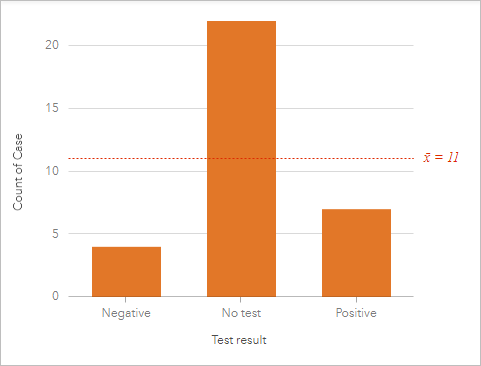


1. Close the Chart statistics pane. Increase the width of the chart card so that all labels are visible.



Next, you will create a second column chart, this time to show test results.

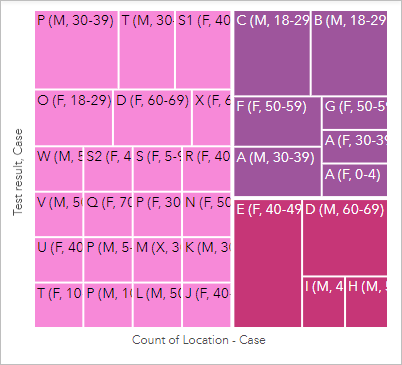
1. In the Case dataset, select Test result, and drag it to the page to create a second column chart. Resize the chart to the same width as the first column chart.



The column chart shows the test results (Negative, No test, and Positive) and the number of cases for each result.

The final step will be to create a chart that shows the test results for each case. You will create the chart using the **joined dataset** so that it will also tell you the number of activities reported for each case.

1. In the data pane, collapse Case and expand Location - Case.
2. Select the Test result and Case fields.
3. Drag the fields to an empty area on the page, point to Chart, then drop the fields on Treemap.

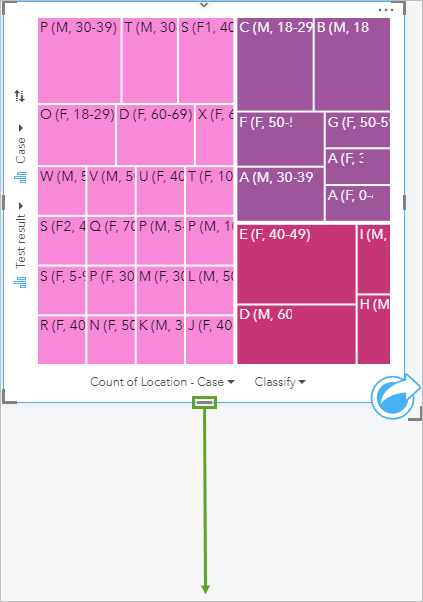


A treemap is created with different sized rectangles for each case. The colour of the rectangle is based on the test result, and the size is based on the number of activities reported for each case.

**Note:**

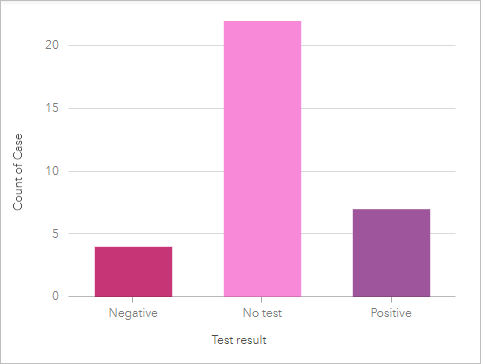
If your treemap does not show the cases grouped by test results, you may have selected the fields in the wrong order in the data pane. Click the Flip fields button on the y-axis to change the orientation of the treemap.

1. Click the treemap to activate it, if necessary, then drag the bottom handle to expand the treemap so the case values are visible.



The treemap and second column chart both show test results. It would be helpful if the colours on the treemap were also used on the column chart.

1. Activate the second column chart (Card 4).
2. Open the Layer options pane for Card 4 and change the Symbol type to Unique symbols.
3. Click Chart statistics and deselect Mean. Close the Chart statistics pane.



The test results on the column chart and the treemap are styled using the same colours.

1. Click Hide on each of the cards to hide the card toolbar and make more space for the charts themselves. Alternatively, you can add descriptive titles to the cards.

Hide button

1. Save your workbook.

Your page is now complete with several column charts, a treemap, and a link chart to help with contact tracing. Next, you will use the charts you made to analyse the viral spread through the community.

**Interpret relationships between cases.**

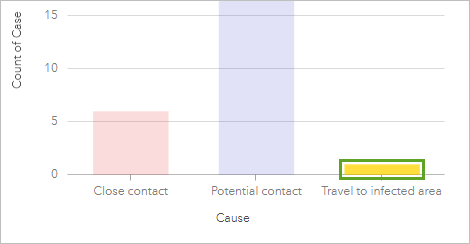
In ArcGIS Insights, the cards on your page are interactive. You will use the interactions between cards to interpret patterns and draw meaningful conclusions from your data.

**Determine viral spread**

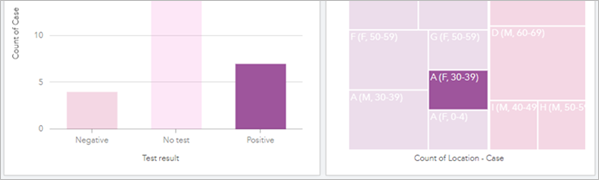
You will use the charts created in the previous topic to determine how the virus has spread, where transmission occurred, and who should be tested or isolated to stop the spread.

The column chart for Cause indicates there is only one case of a known traveller.

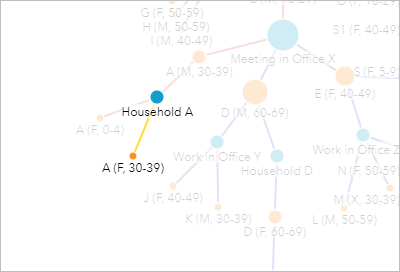
1. In the column chart on Card 3, click the Travel to infected area column.



The other cards on the page update to show the data linked to the selected column on the chart.



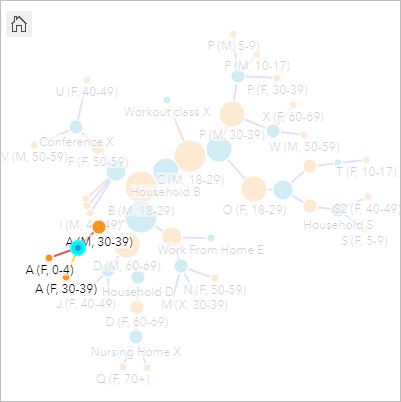
The treemap indicates the case linked to travel was A (F, 30-39), meaning the case was for a female age 30 to 39 in Household A. The treemap and column chart for test results indicate that the test results for this case were positive, meaning the patient was infected.



The **link chart** also shows A (F, 30-39) selected, and the direct connections between the patient and the activities she reported. In the case of A (F, 30-39), only her household was reported.

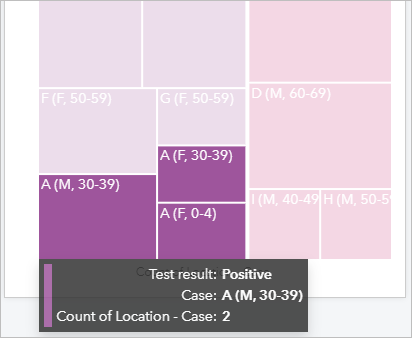
The **reference table** indicates the date that A (F, 30-39) reported her activity (arriving home from travel to Household A) as 3/6/2020. Since A (F, 30-39) did not report any other activity besides returning to her house, she is a low risk for transmitting the virus. You will look at the members of her household next.

1. In the link chart, click the node for Household A to select it.



The link chart shows three cases linked to Household A. The cases are also selected on the treemap. A (F, 30-39) is the known traveller. A (M, 30-39) and A (F, 0-4) are the other members of her household. All three household members have tested positive for the virus.

1. Hover over the three selected rectangles on the treemap.



The hover pop-ups indicate that only one activity was reported for A (F, 30-39) and A (F, 0-4), meaning the only place they visited was their household (in other words, both cases were isolated after exposure to the virus). Case A (M, 30-39) has two reported activities. You will look next at where those activities were located to see whether there are others who may have been infected.

1. On the treemap, click A (M, 30-39) to select it.

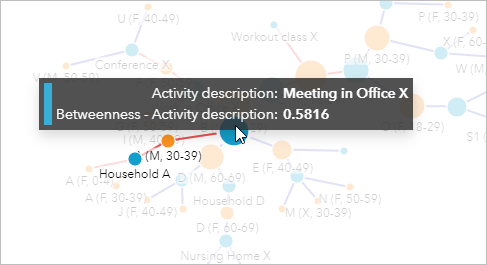
The link chart updates to show the two links connected to A (M, 30-39).

**Note:**

Use the Zoom tools button or your scroll wheel to zoom in on the link chart to see the nodes better.

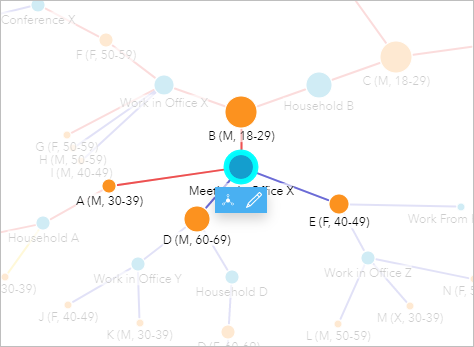
Zoom tool

1. Hover over the two nodes for the activities linked to A (M, 30-39).



One of the nodes is for Household A, which you already explored. The second is for Meeting in Office X. That node is larger than the others you’ve looked at so far, which indicates it is a more important node for connecting nodes to each other (in other words, it has a more important role in spreading the virus).

1. Click the node for Meeting in Office X.



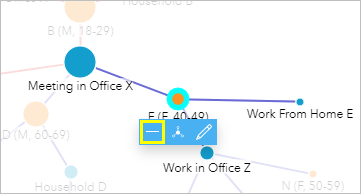
According to the reference table, the meeting took place on 3/9/2020, which is after the time when A (F, 30-39) returned from her trip.

Four cases are connected to Meeting in Office X. Look at the treemap to see more information about the cases. All four have been tested; two tested positive, including A (M, 30-39), and two tested negative. The cases that tested negative do not need to be pursued any further.

1. Click the node for E (F, 40-49) (one of the negative cases).

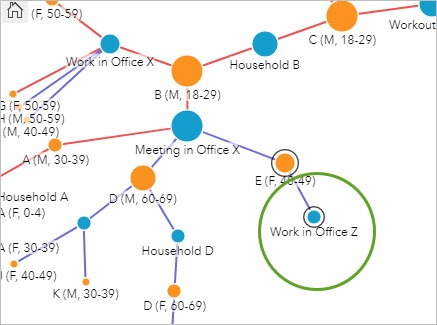
Case E (F, 40-49) reported two activities where she had contacts: working from home and working in office Z. She also reported which coworkers she was in close contact with in office Z. Since E (F, 40-49) tested negative, those coworkers do not need to be tested. You will hide the leaf node in the chart to reduce clutter. When you select a node in the link chart, you can hide its leaf nodes, set it as a central node, or edit it.

1. On the pop-up toolbar, click Hide Leaf Nodes.



The link and node for Work From Home E are hidden.

1. Click the node for Work in Office Z and hide its leaf nodes.

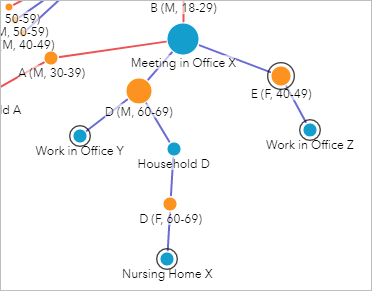


The nodes for the coworkers in Office Z are hidden. The other negative case was D (M, 60-69).

1. Click the node for D (M, 60-69).

D (M, 60-69) reported working in Office Y and staying in Household D. The link to Household D continues to case D (F, 60-69), which in turn leads to Nursing Home X. Since D (M, 60-69) tested negative, none of these cases need to be tested because of their connection to him.

1. Hide the leaf nodes for Work in Office Y and Nursing Home X.

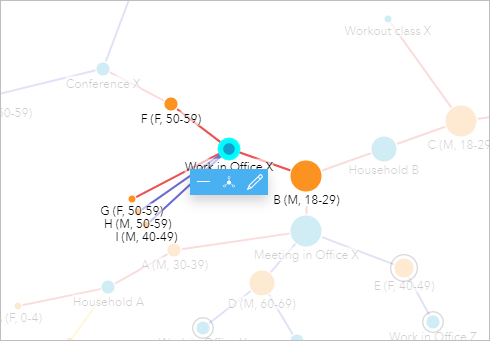


The other case from the meeting in Office X was B (M, 18-29), who tested positive.

1. Click B (M, 18-29) to see his connections. The node has a high betweenness, which indicates this case is important for connecting all the nodes together.

B (M, 18-29) reported his household and Office X as the locations where he was in contact with others.

1. Click Work in Office X to see the information for the coworkers.

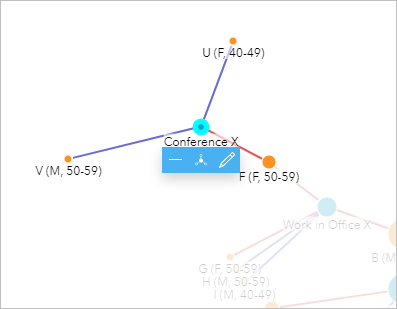


According to the treemap, H (M, 50-59) and I (M, 40-49) both tested negative. Therefore, there is no need to pursue their points of contact any further.

F (F, 50-59) and G (F, 50-59) both tested positive. There is no information for which locations G (F, 50-59) visited since her exposure. That data will need to be collected to fully trace additional points of contact.

F (F, 50-59) reported attending Conference X and coming into close contact with two people.

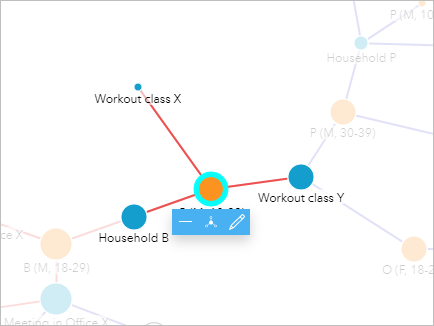
1. In the link chart, click Conference X.



The two conference attendees that came into contact with F (F, 50-59) were U (F, 40-49) and V (M, 50-59); in the treemap, these two cases are listed as No test. Since F (F, 50-59) tested positive, those two cases will need to be **tested as well**. More information will also need to be gathered to determine whether other conference attendees should be tested or directed to self-isolate.

The other location that B (M, 18-29) reported was his household, which is connected to a second person who visited his household.

1. Click the node for C (M, 18-29) (the second node connected to Household B).



C (M, 18-29) is connected to Household B, plus two workout classes. His test results are positive, which means there may be risk to other people who attended the classes.

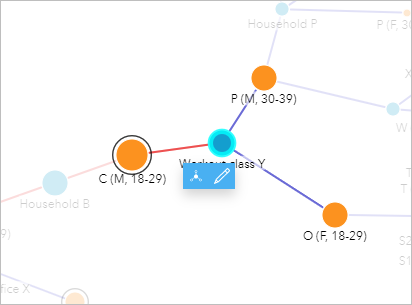
According to the reference table, the activities reported were on three days: 3/8/2020, 3/9/2020, and 3/11/2020. Since C (M, 18-29) likely contracted the virus from B (M, 18-29) on 3/9/2020, the activity on 3/8/2020 (Workout class X) does not need to be pursued further.

Workout class X does not have any other contacts connected to it, so you can hide it.

1. Click Hide leaf nodes.

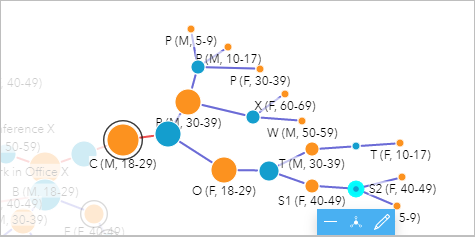
Next, you will examine the second workout class.

1. Click Workout class Y.



Two close contacts were reported, neither of whom have been tested. Since C (M, 18-29) tested positive, tests should be conducted for both cases from the workout class. The two cases also have reported contacts.

1. Hold Ctrl and select the remaining location nodes (Household P, Work in Office α, Workout class Z, Household T, and Household S).



All the connections downstream from Workout class Y are selected. According to the treemap and the column chart with test results, none of the selected cases have been tested. Since there is a known link to a positive test, further action may be required for these cases, such as conducting tests or ordering self-isolation.

1. Click an empty area on the link map to clear the selection.
2. Save your workbook.

You have used charts to visualize and analyse case data to determine which actions were required for the community. Link analysis is a powerful way to analyse relationships in your data and visualize the interconnectedness of the cases that you are analysing. If you have access to patient data and location, such as public health authorities, you can perform the following operations in Insights for mapping the data:

* Aggregate your case data into boundaries using spatial aggregation, such as neighbourhoods or census blocks. Display the total number of cases in each boundary using graduated symbols with the Counts and amounts (size) symbol type or divide the number of cases by the total population of the boundary and display the proportions using graduated colours with the Counts and amounts (colour) symbol type.
* Aggregate your case data into boundaries using spatial aggregation and calculate z-scores to determine distance from the mean. The z-scores can be mapped using standard deviation classification and help determine if there are clusters of high and low values, indicating an outbreak in a certain area or low levels of cases in another.