Organizational Network Analysis

Table of Contents

Install Gephi for Custom Data Visualization

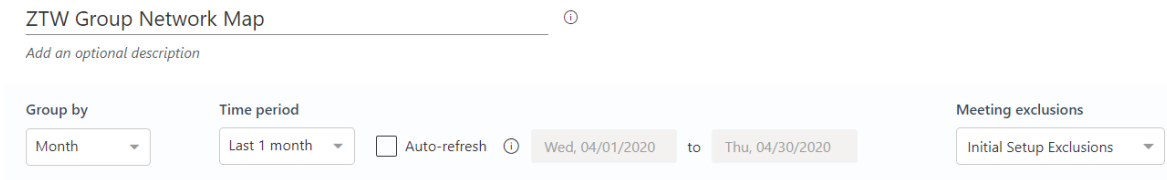
Gephi is an open source network analysis software that enables users to create customized network analysis measures and maps. Gephi is used in this analysis due to node count limitations of dynamically updating the Azure Templates visualizations. Install Gephi using the following steps:

1. Download Java and install – be sure to note the where location Java is installed
2. Download Gephi and install – be sure to note the where location Gephi is installed
   Note: You may have to update the Gephi configuration file with the location for Java according to the following steps:
   a. Locate the folder location that Java is installed. By default, this installs at *C:\Program Files (x86)\Java\jre1.8.0_251*
   b. Open the jre1.8.0_251 folder, then click into the address bar and copy the folder path
   c. Locate the Gephi configuration file (gephi.conf) within the folder where Gephi was installed ("…Gephi-0.9.2\etc"). By default, this installs at *C:\Program Files\Gephi-0.9.2\etc*
   d. Open the Gephi configuration file with a text editor like Notepad
   e. Remove the leading # from the #jdkhome line (towards the end of the Gephi configuration file)
   f. Paste in the Java folder path that you've recently copied to your clipboard so that the Gephi configuration line now reads similar to *jdkhome="C:/Program Files (x86)/Java/jre1.8.0_251"*
   g. Save and close the Gephi configuration file, then open Gephi

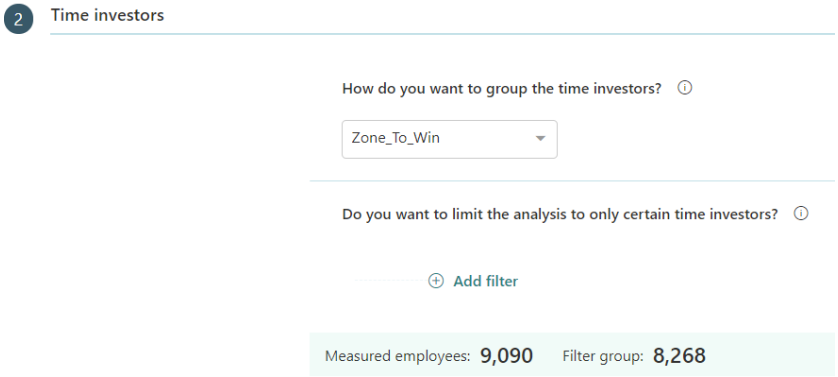
Group-Level Network Map:

Run a Group-to-Group query in WpA

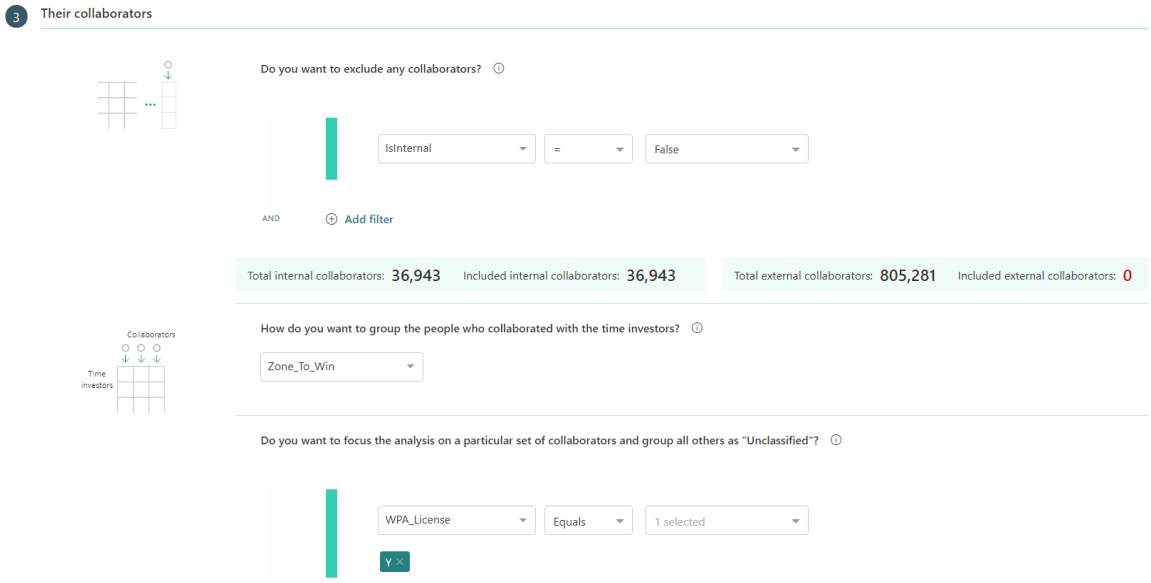
1. Navigate to the Group-to-Group queries page in WpA
2. Setup a query with monthly data aggregation and a date range



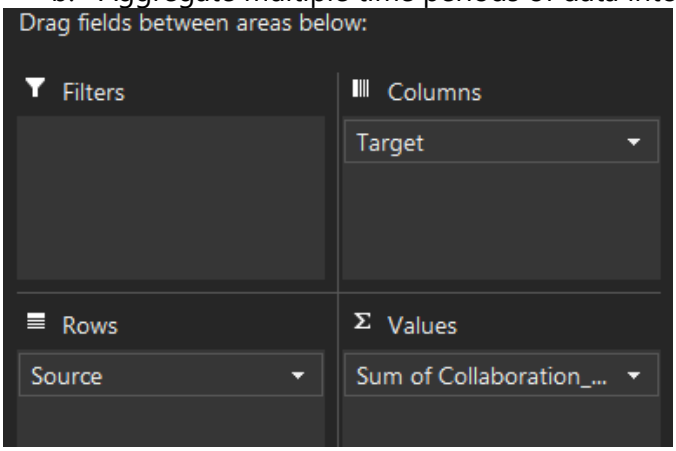
3. Select the metric for Collaboration Hours
4. Group Time Investors by your preferred attribute, Organization,



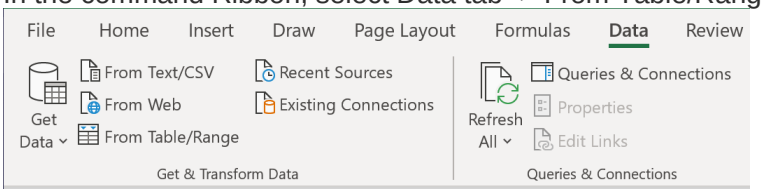
5. Group Collaborators by the same attribute selected in step 4
6. You may choose to exclude external collaborators (IsInternal = False), and focus the analysis on those with a WpA license (WPA_License = Y)



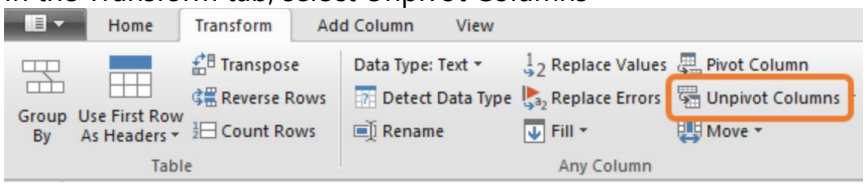
7. Verify your settings and Run your analysis
8. Once your queries have completed, download the csv file
9. Relabel the group names as you would like them to appear on your visual
10. Use a pivot table (Rows=Source, Columns=Target, Values=Collabortaion_hours) to cleanup the following:
    a. Filter out groups that you do not wish to show in your analysis (e.g. "Other_collaborators", groups falling below minimum aggregation size, groups not relevant to this analysis, etc.)
    b. Aggregate multiple time periods of data into a single value



11. Create a new tab and unpivot (manually or automatically through the following steps) your pivot table into a new table with the following 3 columns: Source (from pivot table Rows), Target (from pivot table Columns), and Weight (from pivot table cell values for each row-col combination)
    a. Store data in a table
    b. Select any cell in the table
    c. In the command Ribbon, select Data tab -> From Table/Range



    d. Select the columns you want to Unpivot (selecting the first column and holding the Shift key to select the rest)
    e. In the Transform tab, select Unpivot Columns



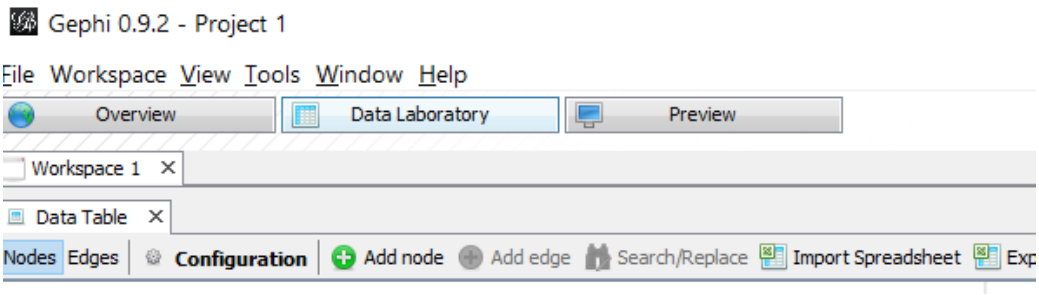
    f. Finally, back to Home tab -> Close & Load
    g. Manually remove some of the rows as needed

12. Save this file to use in the next steps.

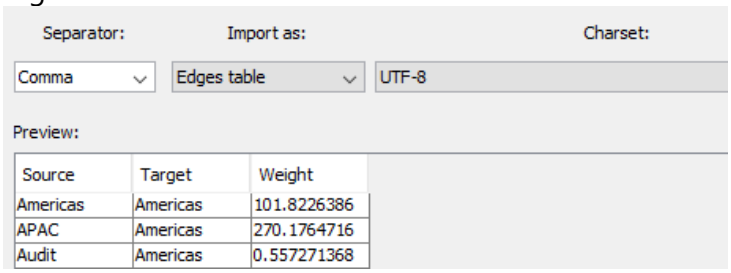
Source	Target	Weight
Americas	Americas	101.8226
APAC	Americas	270.1765
Americas	APAC	0.557271
APAC	APAC	3.335422

Create your Group-Level Network Map

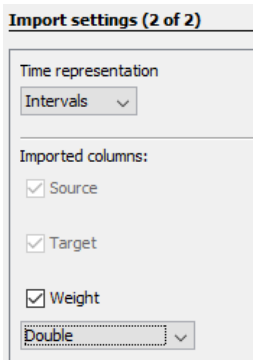
1. Open Gephi and create a New Project
2. Click on the Data Laboratory button and click Import Spreadsheet



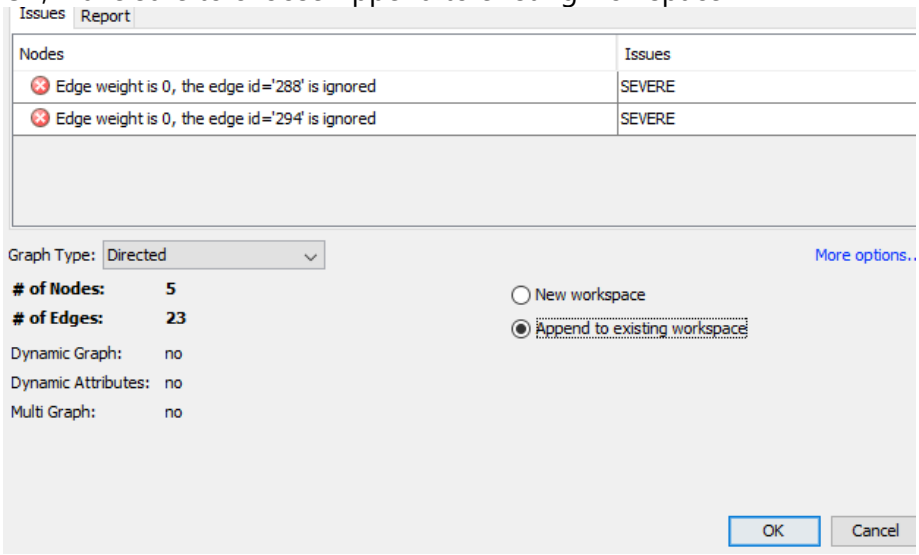
3. Navigate to the folder where you saved your Group-to-Group query and select the file you created in previous section. This will serve as the underlying data for your network map.
4. Select the tab with the cleansed data from the file you prepared and verify that Gephi recognizes this file as an Edges Table. If not, select this value from the dropdown
   Note: Sometimes Gephi does not recognize you Source and Target column due to a problem with the UTF-8 encoding. If you have this problem, try saving the file as a regular csv.



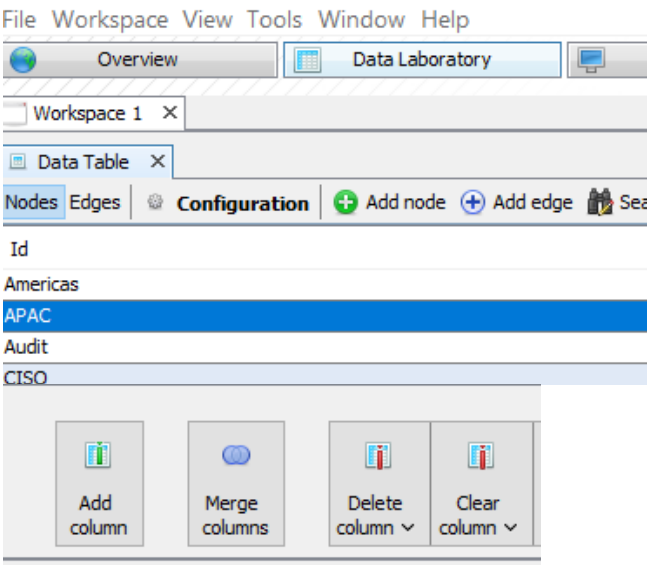
5. Click next and make sure that Gephi recognizes your Weight column as a Double data type, the hit Finish



6. You can likely ignore the errors, especially those about Edge weight is 0. Before hitting OK, make sure to choose Append to existing workspace



7. To add groupings for coloring, you'll need to add additional attributes to your nodes.
8. You can do it manually or by creating a Node file:
   Manual process within Gephi tool is a good option if you have limited number of rows. Uploading the Edge file, Gephi automatically detects the Nodes.
   a. Click the Nodes tab and Select Add column from the bottom of the page.

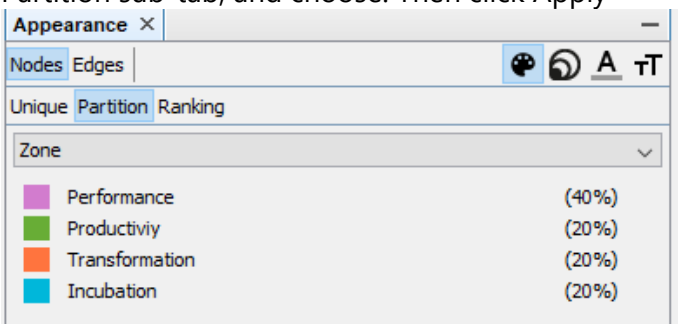


b. Title your new column and select Data Type = String
c. A new column will appear in your Nodes table. Double click the first row and add the info that this group falls under. Repeat this for other attributes you want to add about each group (e.g. GroupSize). Additionally, you can choose input custom labels for each node in the Label column.

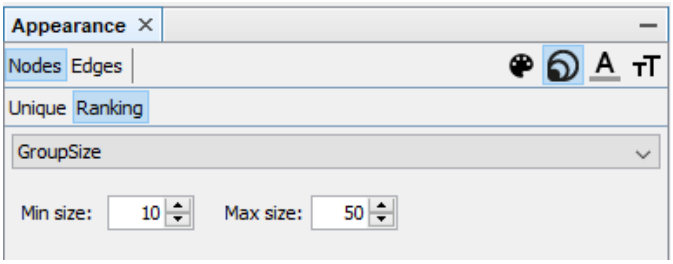
Prepare a Node file with your desired columns

a. Id column is required, and it should correspond to the groups you have as source and target columns in the Edge file
b. You can then add additional columns like Label, GroupSize, Organization etc. These columns can later be used for coloring and sizing of the nodes in the visualization.

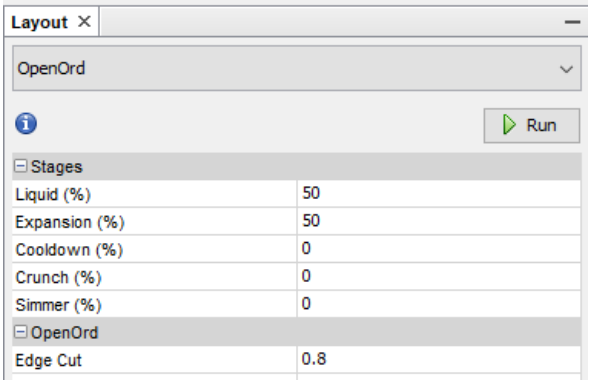
9. Navigate back to the Overview Tab at the top of the page to see an initial view of your group-level network
10. To color your nodes by an attribute, click the palette icon from the Nodes tab, click the Partition sub-tab, and choose. Then click Apply



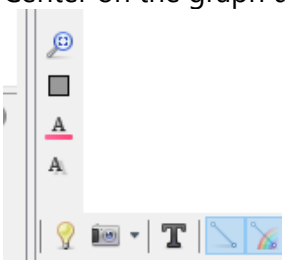
11. Size your nodes by clicking on the Circles icon and selecting an attribute.



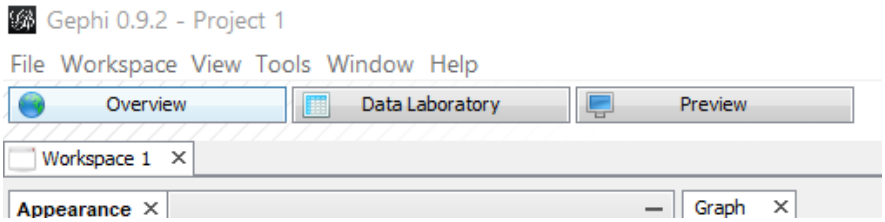
12. You can also color your Edges (Collaboration weights between each node)
13. Click on the overview tab to begin creating your visual and run through the following procedure to tune your visual in the Layout pane on the left
    a. Select OpenORD and run after inputting the following settings
        i. Liquid = 50
        ii. Expansion = 50
        iii. Other settings in the Stages section = 0
        iv. Keep defaults for the Open Ord Settings



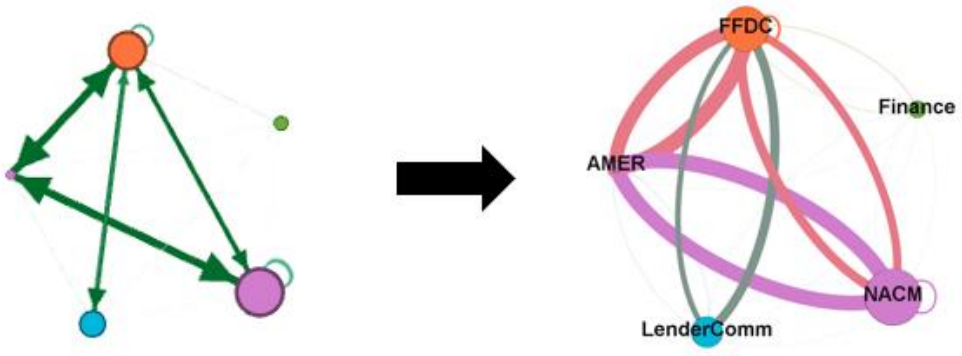
    b. Next select Force Atlas 2 and run after inputting the following settings
        i. Scaling =500
        ii. Gravity = 0
        iii. Make sure dissuade hubs is checked
        iv. Keep other parameters at default
    c. Stop Force Atlas 2 after the nodes are sufficiently spread out
    d. Change the Scaling parameter to 1.0 (instead of 500 or 1000) and run again until the node movement is significantly slowed down
    e. Center on the graph using the magnifying glass icon



14. Once you have finished setting up your network visualization, click into the Preview page to polish you network graph



15. In the Preview Settings panel, adjust Borders, Labels, Line Thickness, Font, etc. to polish your network. See below for the difference between what we setup in the Overview page vs the polishing in the Preview page



Individual-Level Network Map:

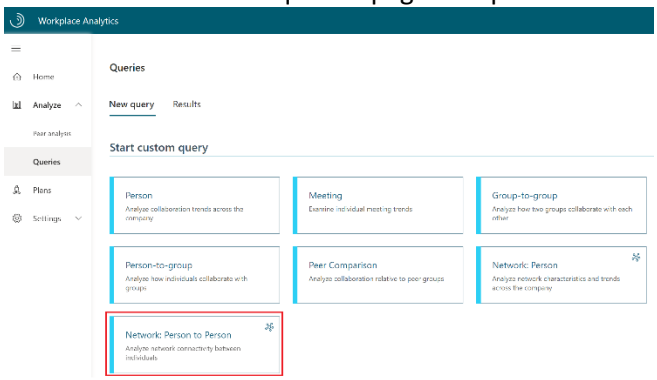
Create a person-to-person interaction table:

You have a few options to create a person level interaction table.

- WPA **Network: Person-to-Person** Query
- WPA Person-to-Person query via **Group-to-Group** query
- WPA Azure Templates interaction table

WPA Network: Person-to-Person Query

1. Navigate to the Network: Person-to-Person queries page in WpA.



2. Setup a query with your desired data aggregation level and a date range.
3. Select the "Strong and Diverse tie scores" metric and filters. Filters are not required unless there is a specific group you would like to focus on in the analysis.
4. Run the query and export the results once the run is complete.
5. Select the following columns, rename accordingly.
    a. TieOrigin_zId -> Source
    b. TieDestination_zId -> Target
    c. StrongTieScore -> Weight
6. Copy the Source, Target and Weight columns over to a new tab and save this tab as a csv named **CleansedInteractions.csv**

    Note: For Source and Target columns, you can also use TieOrigin_PersonId and TieDestination_PersonId columns but for consistency with some of the other methods of creating an interaction table we have selected the zId columns.

7. Follow the steps in section WPA Person Query to create a node file to create a node file to go with the interaction table you just created.

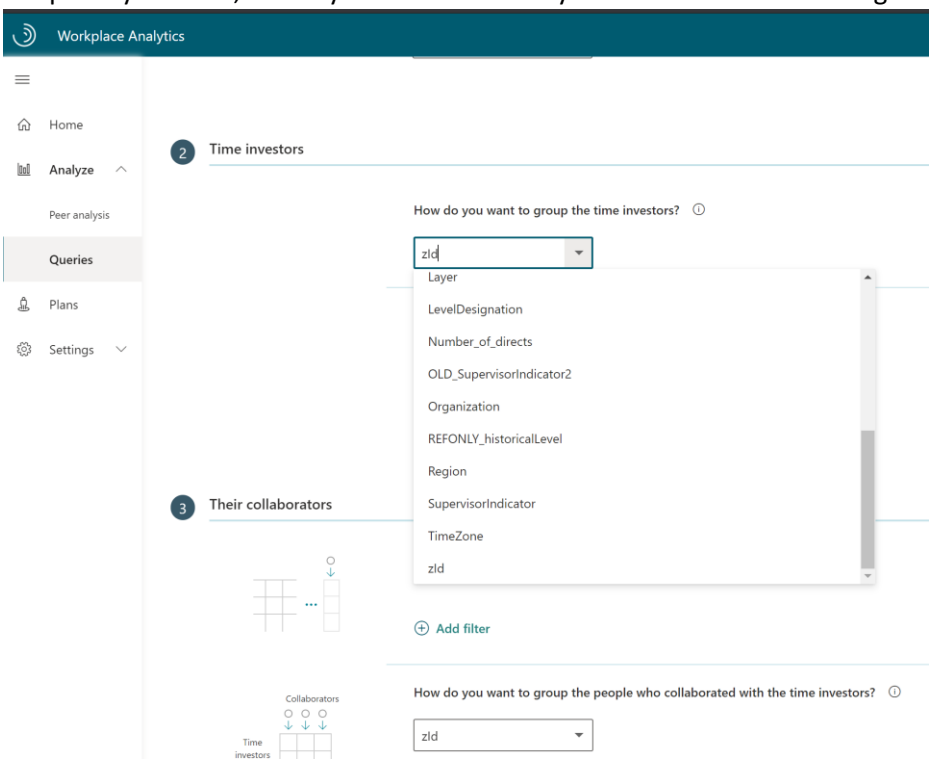
WPA Person-to-Person query via Group-to-Group query

WPA does not have a Person-to-Person query as an out-of-the-box query template. However, with a hack you can turn a Group-to-Group query into a Person-to-Person Query.

PersonId, the main person identifier in the WPA Organizational data, will be hashed (de-identified) in your reports and not available to be selected as a your group collaborator in the Group-to-Group query.

However, if you add an additional person identifier to your HR Org file when uploading (generate a column that has anonymized pids for each person), you can use that metric in **Time Investors** and **Their Collaborator** of the Group-to-Group query to change the group collaboration to a person level collaboration. In the screenshot below **Zid** is that secondary id.

- Important Note: Please make sure the additional id (**Zid**) is not your employee id or any identifiable number that would allow the analyst to identify and analyze the interactions at the person level.
  For privacy reasons, WPA by default de-identify the PersonId when HR org file is uploaded.



To create the interaction matrix (or Edge list) for the ONA analysis, open a Group-to-Group query

1- Navigate to the Group-to-Group query page in WpA.
2- Setup a query with your desired data aggregation level and time period.
3- Select the metrics like Collaboration hours, Email hours, Meeting hours and set filters as needed.
4- Select the Zid in the "**Their collaborators**" and "**Organizational data**" sections of the Person-to-Group query.
5- Run the query and export the results once the run is complete.
6- Select the following columns, rename accordingly.
   a. TimeInvestors_zId (zId) -> Source
   b. Collaborators_zId -> Target
   c. Collaboration_hours (or any other metric that you prefer to be used) -> Weight
7- Copy the Source, Target and Weight columns over to a new tab and save this tab as a csv named **CleansedInteractions.csv**
8- Follow the steps in section WPA Person Query to create a node file to create a node file to go with the interaction table you just created.

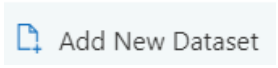
Note: With the same process of introducing an additional id (zId), you can transform the WPA Person-to-Group query to a Person-to-Person query in case the metrics available in the Person-to-Group query (like Email counts) is of your interest as part of the interaction table. In this case, the selection and renaming of the columns should go as follows:

a. TimeInvestors_zId (zId) -> Source
b. Collaborators_zId -> Target
c. Collaboration_hours (or any other metric that you prefer to be used) -> Weight

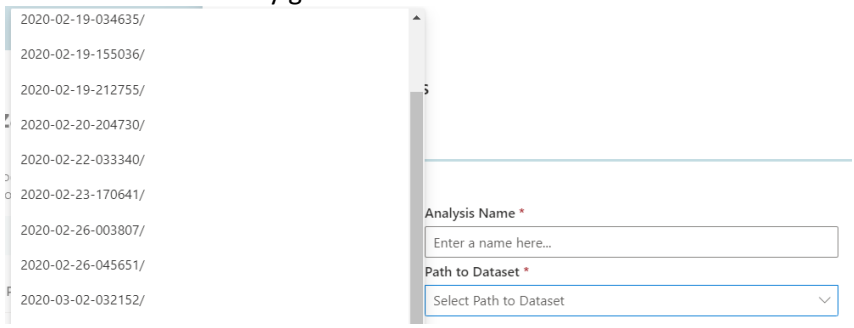
WPA Azure Templates interaction table

Generate a Dataset for Analysis through Azure templates:

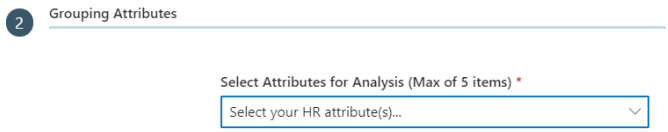
1- Login to your Azure Templates environment using your WpA credentials
2- Click into the Organizational Networks Analysis Template
3- Add a new dataset by clicking at the upper right corner of the table

4- 

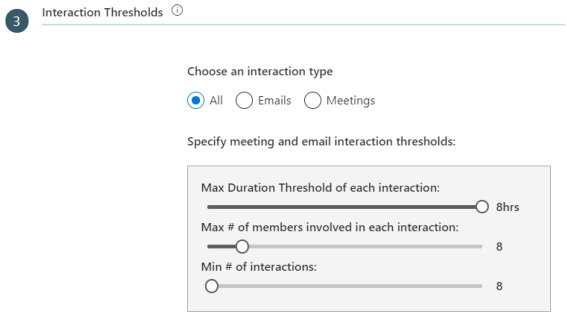
5- Name your dataset in the Analysis Name input box
6- Select the path to the most recently generated dataset based on the file names



7- Choose up to 5 Org Data attributes to use as groupings in your analysis.
8- Note: It's best to only choose the groups that you will use in the. Adding additional groups will bloat the files that you work with later.

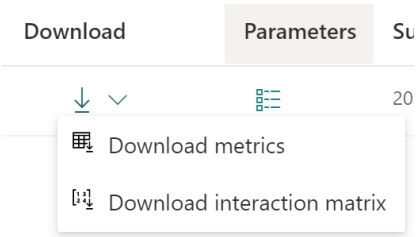


9- Set the interaction thresholds to determine which individuals are meaningfully connected to each other. For the standard analysis, we use the below settings.



Note:

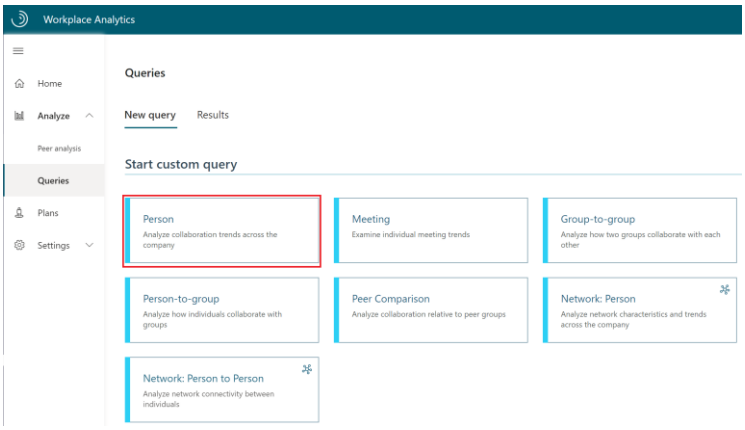
- *Max Duration Threshold notes the duration of the longest meetings that will be considered as a connection point for individuals. Any meetings of duration longer than your setting from this analysis will be excluded.*
- *Max # members refer to the maximum number of participants that can be included in an event that we still consider to be a meaningful connection point. E.g. An all-hands meeting with 5k people would not create a meaningful connection point between all 5k people.*
- *Min # of interactions refers the minimum number of shared meaningful connection points that we want to look for before we assume that 2 individuals are meaningfully connected. E.g. A single email to 3 people would not automatically bring all 3 individuals into one's network.*

1. Add the meeting exclusion keywords you wish to use.
2. Submit your dataset for processing – this will take a while to process.
3. Once the dataset status is checked as complete, click on the download arrow and download the interaction matrix which is your person level interaction matrix and then the metrics.
4. 

5. Open the **interaction matrix** file and rename columns as following:
    a. Node1Pid -> Source
    b. Node2Pid -> Target
    c. WeightbyHours -> Weight
6. If there are multiple months of data, filter your dataset down to a single month or aggregate metrics across all dates into a single time period for analysis
7. Copy the Source, Target and Weight columns over to a new tab and save this tab as a csv named **CleansedInteractions.csv**
8. Open the person level interaction matrix (called Node_Level_Metrics.csv within the downloaded metrics folder)
9. Delete any Org Data columns that you don't plan to use for grouping or sorting. Also delete Column A and the phid column to reduce the file size
   Note: Make sure to leave the Network Analysis metrics at the far right of the table
10. Filter and aggregate the data so that you get a single value for each employee (pid)
11. Delete the date columns and rename the column pid -> ID. Then save the Node Level Metrics tab as its own csv named **CleansedNodeMetrics.csv** and close the file.

WPA Person Query to create a node file

1- Navigate to the Person query page in WpA.

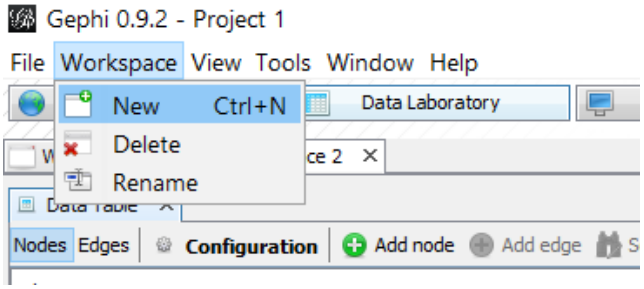


2- Setup a query with the following settings
   a. Group by -> "Month"
   b. Time Period -> Last 1 month"
   c. Select metrics -> minimum one metric but it is not important for our purpose here.
   d. Select filters -> All employees or Active employees.
   e. Organizational Data -> **zId** in addition to any organizational data that plan to use for grouping, sorting, coloring and sizing of the nodes.

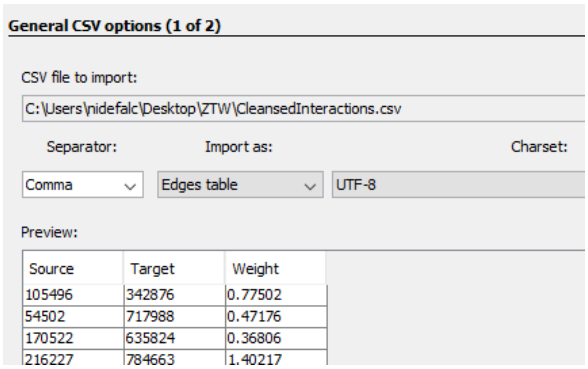
3- Run the query and export the results once the run is complete.
4- Rename the column zId -> ID.
5- Filter and aggregate the data so that you get a single value for each employee (ID aka zId)
6- Select the ID column and the organizational data columns (i.e. Organization, FunctionType) you selected in step 2e, copy into a new tab and then save the tab as its own csv named **CleansedNodeMetrics.csv** and close the file.
7- You can also filter the node file to only include the nodes/employees/ IDs that are either a Source or Target in the **CleansedInteractions.csv** file you just created.

Create your Person-Level Network Map

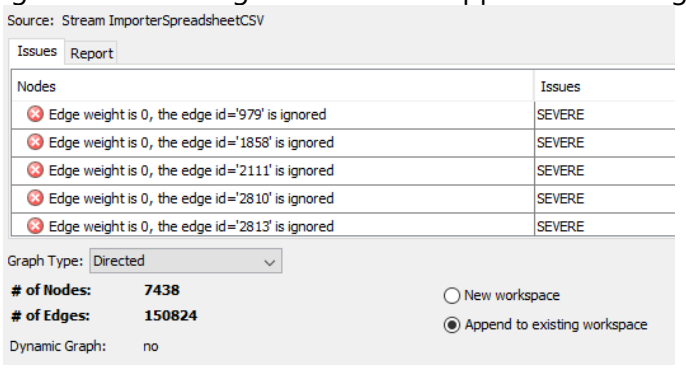
1. Create a new workspace in Gephi by going to Workspace >> New, then clicking on the Data Laboratory button



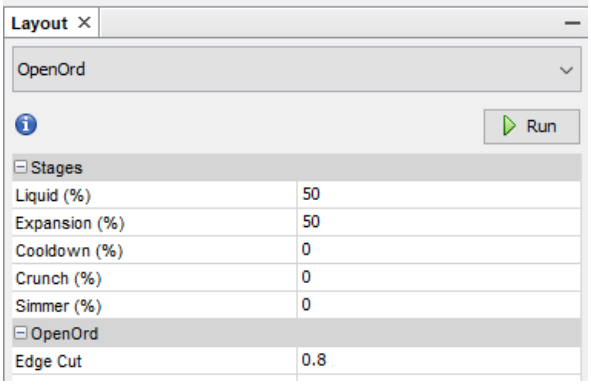
2. Click Import Spreadsheet and select your **CleansedInteractions.csv** file that you created via one of the approaches above. This is the underlying data for your network map.
3. Make sure that Gephi recognizes this file as an Edges table and click Next in the wizard.



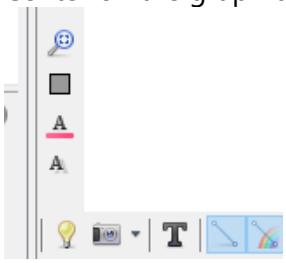
4. Make sure the Weight column is identified as a Double data type and then click Finish.
5. Ignore the warnings and select to Append to existing workspace.



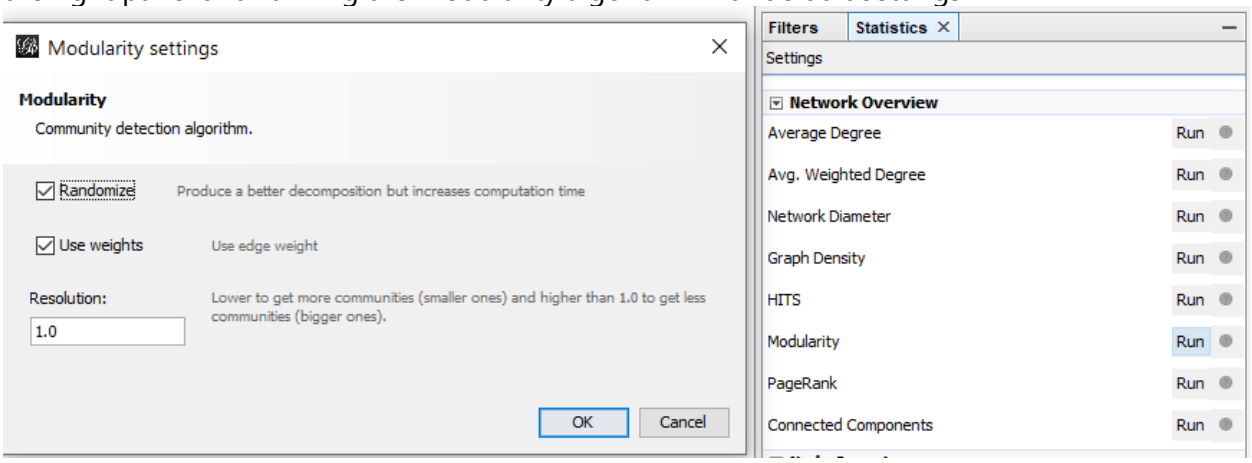
6. Click Import Spreadsheet again and select your **CleansedNodeMetrics.csv** file. This additional data will enable grouping, coloring and node sizing based on your Org Data (and the Azure Templates computed metrics if you are using that approach)
   Note, if you get a Gephi memory error, follow the instructions at this link to increase the memory allotment https://gephi.org/users/install/#memory
7. This time, make sure that Gephi recognizes this as a Nodes table. If not, change using the dropdown, then click Next
8. Make sure that Gephi detects the right datatypes for each column, the click Finish.
9. Choose to Append to existing workspace and click OK.
10. Click on the overview tab to begin creating your visual and run through the following procedure to tune your visual in the Layout pane on the left.
    a. Select OpenORD and run after inputting the following settings
        i. Liquid = 50
        ii. Expansion = 50
        iii. Other settings in the Stages section = 0
        iv. Keep defaults for the Open Ord Settings



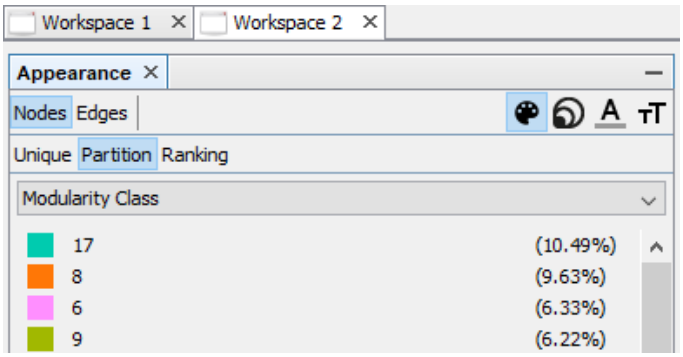
b. Next select Force Atlas 2 and run after inputting the following settings
   i. Scaling =500
   ii. Gravity = 0
   iii. Make sure dissuade hubs is checked
   iv. Keep other parameters at default
c. Stop Force Atlas 2 after the nodes are sufficiently spread out
d. Change the Scaling parameter to 1.0 (instead of 500 or 1000) and run again until the node movement is significantly slowed down
e. Center on the graph using the magnifying glass icon



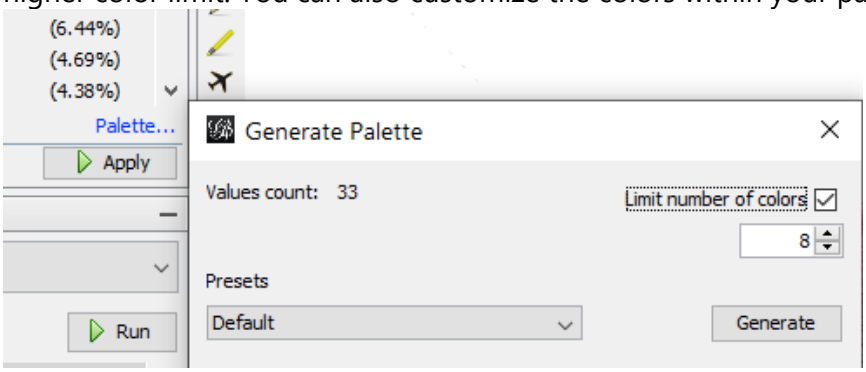
11. You can also create the organic community groupings by selecting the statistics tab on the right panel and running the Modularity algorithm with default settings



12. Color your network map by modularity grouping by clicking on the Nodes tab in the Appearance panel, selecting the Partition subtab, and choosing the Modularity Class grouping as your color designation. Hit Apply.



13. You can increase the number of groups that are colored by clicking on the Palette button at the bottom right of the Appearance panel and choosing to Generate a new palette a higher color limit. You can also customize the colors within your palette.



14. Adjust the bubble sizes using any of the nodel level metrics you imported from Azure Templates if you'd like.