

## Visualizing Relationships and Connections in Complex Data Using Network Diagrams in SAS® Visual Analytics

Stephen Overton, Ben Zenick, Zencos Consulting

### ABSTRACT

Network diagrams in SAS® Visual Analytics help highlight relationships in complex data by enabling users to visually correlate entire populations of values based on how they relate to one another. Network diagrams are appealing because they enable an analyst to visualize large volumes and relationships of data and to assign multiple roles to represent key factors for analysis such as node size and color and linkage size and color. SAS Visual Analytics can overlay a network diagram on top of a spatial geographic map for an even more appealing visualization. This paper focuses specifically on how to prepare data for network diagrams and how to build network diagrams in SAS Visual Analytics. This paper provides two real-world examples illustrating how to visualize users and groups from SAS® metadata and how banks can visualize transaction flow using network diagrams.

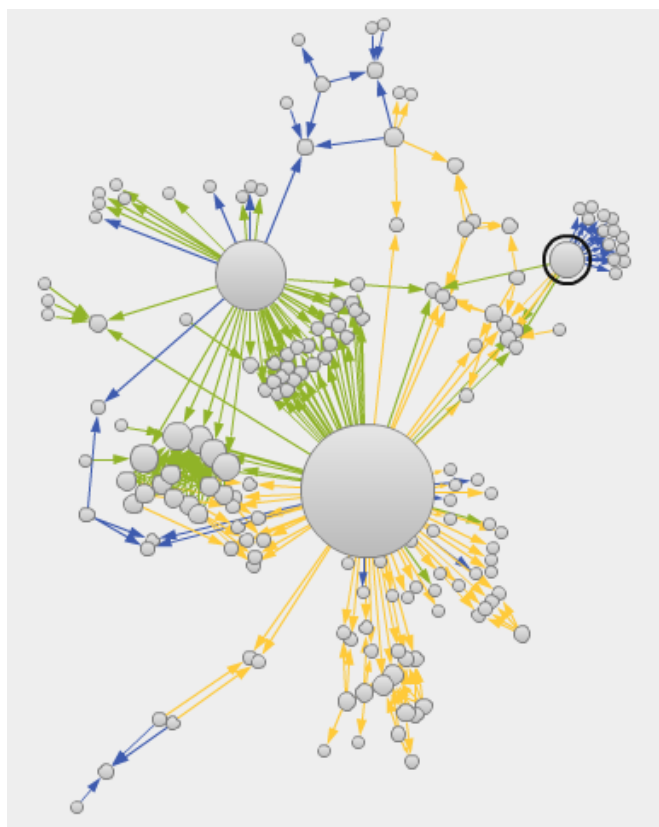
### INTRODUCTION

Network diagrams became widely popular when social media outlets opened their doors for developers to create visual representations of a person's social network. Users of Facebook and LinkedIn can visualize their entire network of connections using third party developer tools that utilize network diagrams to mathematically group and correlate relationships to highlight mutual connections and visualize clusters of relationships.

A network diagram is composed of nodes and links. Each link represents a relationship between two values represented as nodes. The effectiveness of a network diagram can be defined by the sizing, coloring, spacing, and clustering of nodes and links. Lucky for us, SAS Visual Analytics handles all of the hard work. Figure 1 shows an example network diagram visualizing relationships between users and groups within SAS Metadata.

This paper provides instructions for building network diagrams in SAS Visual Analytics Explorer. This paper also provides two real world examples of how network diagrams are useful. Readers of this paper should have a basic understanding of SAS Visual Analytics Explorer and how to prepare data for analysis in SAS Visual Analytics.

Note: This paper is based on SAS Visual Analytics version 7.1.



**Figure 1: Sample Network Diagram output from SAS Visual Analytics Explorer.**

## UNDERSTANDING DATA FOR NETWORK DIAGRAMS IN SAS VISUAL ANALYTICS

In the world of analytics, everything starts and ends with the data. Network diagrams require that the data is structured with either hierarchical or ungrouped node-link pairs. If this requirement is met, network diagrams can easily be generated and designed to provide intuitive analysis.

There are two types of networks in a network diagram:

- **Hierarchical** – uses a standard hierarchy structure of categorical values.
- **Ungrouped** – uses a traditional node-link, or source and target data values.

In most cases, data structures will fall into one network type easier than the other. Figure 2 shows how to select the network type in SAS Visual Analytics Explorer.

### HIERARCHICAL NETWORK

The hierarchical network type is based on a set of categorical variables that make up a classic hierarchy. This is the most common data format and easiest to utilize. Traditional hierarchies follow a logical path down a related set of fields. For example, a locational hierarchy may contain only geographical attributes like country, state, city, and zip code. Due to the nature of network diagrams and visual capabilities, traditional hierarchies can expand and combine unrelated attributes to form complex networks.

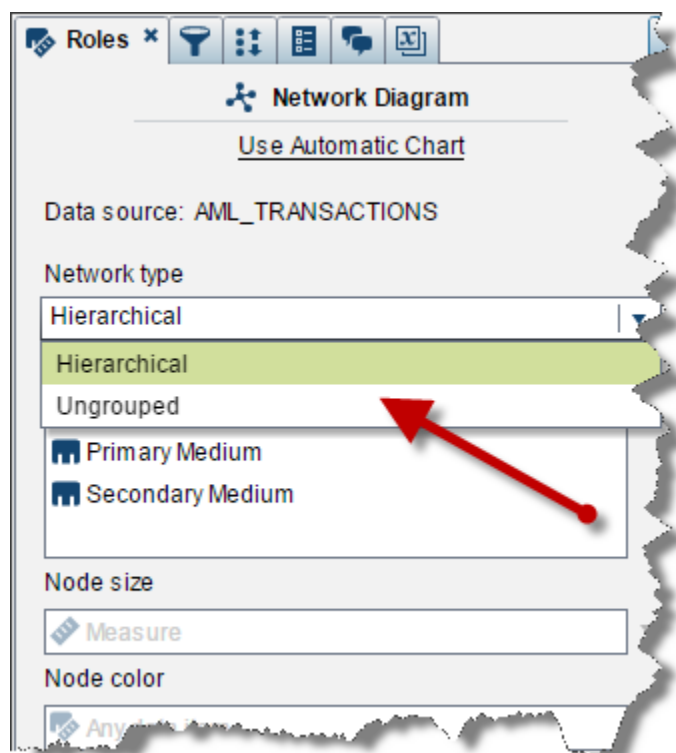


Figure 2: Selecting Network type in SAS Visual Analytics Explorer.

A basic hierarchical network diagram must contain, at a minimum, a 2-level hierarchy. This paper provides a real world example, which uses a 2-level hierarchy for simplicity. The nature and complexity of the data will drive how deep (i.e. the number of levels) hierarchies will go. Hierarchies traditionally came from OLAP “drill down” functionality through a particular subject area within the data. However, network diagram hierarchies expand beyond drilling down or into data by adding visual context to the relationships between members of the levels within a hierarchy. Think of a hierarchical network graph as a fully expanded hierarchy of values showing the relationship of all possible members within a given hierarchy. Figure 3 shows an example of hierarchical data.

Transaction CDI ▲	Primary Medium	Secondary Medium	Amount
DEBIT	DEBIT CARD	PIN	\$225.76
DEBIT	CASH EQUIVALENT	TRAVELERS CHECK	\$616.42
DEBIT	WIRE	INTERNATIONAL	\$374.66
DEBIT	CASH	N/A	\$78.29
DEBIT	EXTERNAL TRANSFER	ACH	\$369.17
DEBIT	INTERNAL TRANSFER	ACH	\$11.10

Figure 3: Example of hierarchical data for a network diagram.

## UNGROUPEd NETWORK

An ungrouped network consists of “source” and “target” values that represent the relationships or connections between nodes. SAS Visual Analytics Explorer visualizes ungrouped networks by displaying a node for each value of a source data item, then linking another node corresponding to the target value. One way to think of the structure of ungrouped data is to consider records as conceptual objects that have parent associations. All possible values should exist as objects, even parent nodes that do not have any additional connections. These are referred to as terminal objects because they have no target or parent node to reference. Figure 4 shows an example of ungrouped data. Records with no parent values are terminal records. These are necessary to complete an ungrouped network diagram.

	object_id	object_displayname	parent_id	parent_displayname
1	A5TYBQPW.A5000003	SAS Administrators	A5TYBQPW.A500000X	Theme Designer for Flex: Administration
2	A5TYBQPW.A5000003	SAS Administrators	A5TYBQPW.A500000N	Management Console: Advanced
3	A5TYBQPW.A5000003	SAS Administrators	A5TYBQPW.A5000006	Metadata Server: Operation
4	A5TYBQPW.A5000003	SAS Administrators	A5TYBQPW.A5000005	Metadata Server: User Administration
5	A5TYBQPW.AP000001	SAS Administrator	A5TYBQPW.A5000003	SAS Administrators
6	A5TYBQPW.A50000UV	[A] Super Developer	A5TYBQPW.A5000003	SAS Administrators
7	A5TYBQPW.AP000006	SAS Environment Manager Service Account	A5TYBQPW.A5000003	SAS Administrators
8	A5TYBQPW.A5000003	SAS Administrators		

Figure 4: Example of ungrouped data for a network diagram.

## HOW TO BUILD A NETWORK DIAGRAM IN SAS VISUAL ANALYTICS EXPLORER

Like many visualizations in SAS Visual Analytics, it is very easy to produce a network diagram. Before building a network diagram, the data must fit into the roles required in order to produce the network diagram. In other words, the structure of source data must contain valid columns in order to be represented on a network diagram.

To create a network diagram, select “Network Diagram” from either of the Visualization menus shown in Figure 5 below.

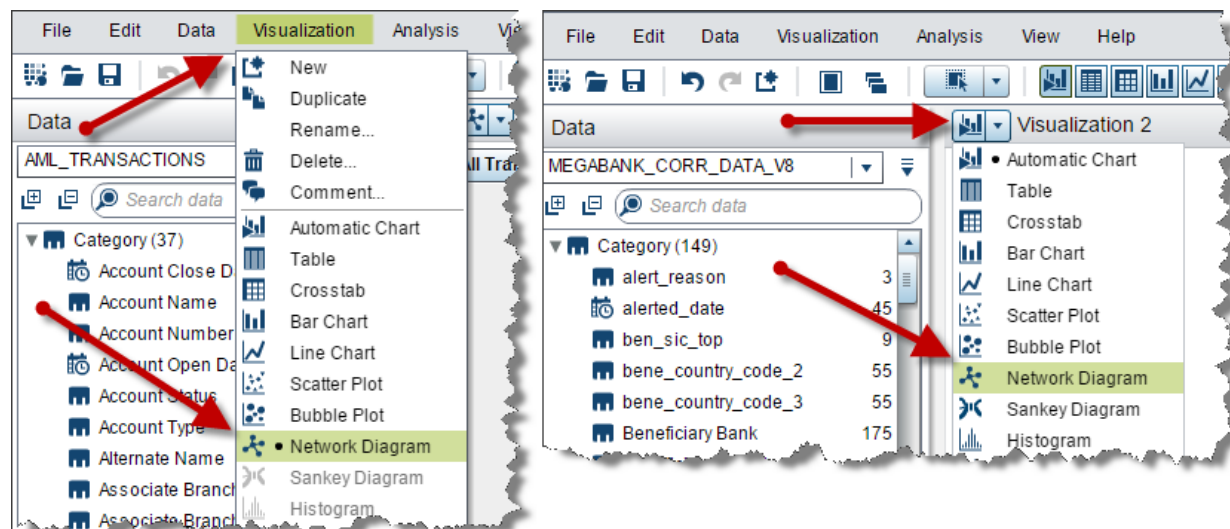


Figure 5: Creating a Network Diagram in SAS Visual Analytics Explorer.

As described earlier, begin your analysis by understanding the source data and select the proper network type to display using the Roles tab on the right.

## NETWORK DIAGRAM ROLES

Drag and drop data items from the left into respective roles on the right. Hierarchical network types have different roles than ungrouped network types. Use categorical hierarchy levels in the Levels role to build hierarchical networks. Use source and target data items in the respective Source and Target roles to build ungrouped networks.

In addition to building the basic network, you can also specify node size and color as well as link size and color. These are especially useful for highlighting key facts about nodes and relationships of nodes using discrete or numerical values. Data tips are useful for reference when hovering over nodes in the network diagram.

- **Node size** – useful for representing differences in measures such as dollar amounts, volumes, or other numeric attributes.
- **Node color** – useful for representing differences in numeric or character-based values that represent intensity or call attention to specific nodes.
- **Link size** – similar to node size, link size is useful for representing numeric measures that represent dollar amounts, volumes, or any numeric measure, which represents the strength of a relationship.
- **Link color** – similar to node color, link color is useful for numeric measures or character-based classifications of the relationships between nodes.

## NETWORK DIAGRAM PROPERTIES

Adjust the properties of network diagrams to customize the look and feel of the visualization. Geomaps are especially useful if the source data contains geographical values for the data items used in the network diagram. To enable the geomap underlay, select the Show map option on the network diagram properties as shown in Figure 6. Be sure to define the data items as geographical items to enable the geomap feature.

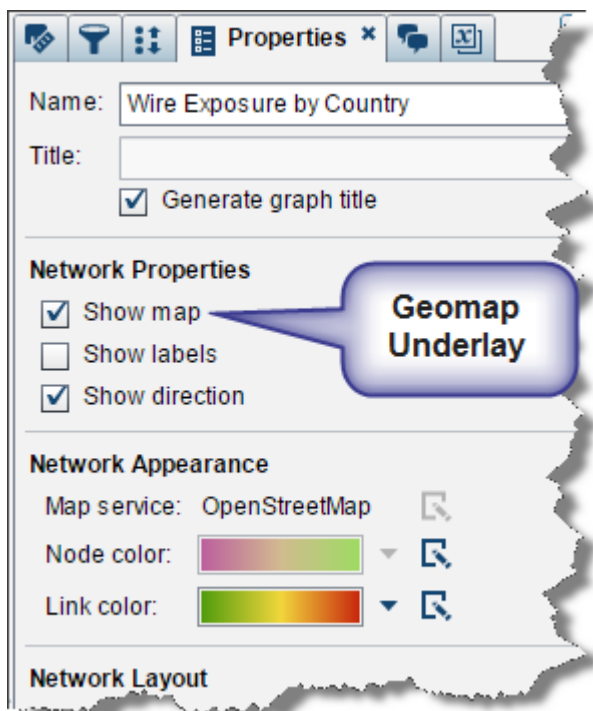


Figure 6: Enabling geo maps in network diagrams.

Other useful properties can be adjusted such as direction of the node links, color gradients, and other textual references in the network diagram. Refer to the SAS Visual Analytics User Guide for additional details on arranging nodes on the diagram and other useful tips.

## EXAMPLE: VISUALIZING USERS, GROUPS, AND ROLES FROM SAS METADATA

The SAS Metadata server is a core component of the SAS platform and is used to manage user access and functionality within a SAS environment. Users are created and managed using the User Manager of SAS Management Console. Users can be members of groups or roles. Groups and roles can be nested as members of other groups and roles. Groups can be used to organize access to SAS metadata folders, solution-specific functionality, or to simplify the management of users and other groups. Roles are specifically used to manage actions or functionality a user can perform within a SAS environment.

### RETRIEVING IDENTITY INFORMATION USING THE %MDUEXTR MACRO

Users, groups, and roles are perfect for ungrouped network diagrams. In order to extract this information, the %MDUEXTR macro is used to collect data from the SAS Metadata server. Output data from the %MDUEXTR macro must be formatted and restructured into a single dataset. The goal is to stage a dataset with each record representing a metadata object (user, group, or role), as well as a target parent link to the respective group the metadata object belongs to.

```
/* Extract all user/group/role metadata - requires administrator user */
%mduextr(libref=work);
```

The %MDUEXTR macro is provided with a SAS installation and by default is available in any session as a compiled macro. It has to be run by a user authorized as an administrator of the respective environment. Output datasets of this macro provide information about users, groups, and roles as configured in the User Manager of SAS Management Console. The GROUPMEMGROUP\_INFO table provides group memberships within other groups. The “id” columns represent the parent group, while the “mem” columns provide the groups, which are a member of the respective parent group. Additional attributes about the group and parent group are joined using the IDGRPS table.

The following code stages group and role relationships, and joins additional details:

```
proc sql;
  create table gginfo as
  select
    groupmemgroups_info.memId as object_id,
    groupmemgroups_info.memName as object_name,
    object.displayname as object_displayname,
    coalesceC(upcase(object.grpType), 'GROUP') length=20 as object_type,
    object.description as object_description,
    groupmemgroups_info.id as parent_id,
    groupmemgroups_info.Name as parent_name,
    parent.displayname as parent_displayname,
    coalesceC(upcase(parent.grpType), 'GROUP') length=20 as parent_type,
    parent.description as parent_description
  from groupmemgroups_info
    left join idgrps as parent on groupmemgroups_info.id = parent.keyid
    left join idgrps as object on groupmemgroups_info.memid = object.keyid
  ;
quit;
```

The following code stages user memberships for groups and roles, and also joins in additional details:

```
proc sql;
  create table gpinfo as
  select
```

```

    groupmempersons_info.memId as object_id,
    groupmempersons_info.memName as object_name,
    usr.displayname as object_displayname,
    'USER' length=20 as object_type,
    groupmempersons_info.memDesc as object_description,
    usr.title as user_title,
    groupmempersons_info.id as parent_id,
    groupmempersons_info.name as parent_name,
    group.displayname as parent_displayname,
    coalesceC(upcase(group.grpType), 'GROUP') length=20 as parent_type,
    group.description as parent_description
from groupmempersons_info
    left join idgrps as group on groupmempersons_info.id = group.objid
    left join person as usr on groupmempersons_info.memid = usr.objid
;
quit;

```

After user, group, and role memberships are extracted they are appended together along with the IDGRPS source table. The IDGRPS table contains a record for every group and role. Appending the IDGRPS table ensures groups and roles can terminate properly in an ungrouped network type.

The following code appends necessary data staged previously, cleans up variables, and defines a relationship type variable:

```

data sgf2015.metadata_user_object_rels;
length
    object_id $20 object_name $100 object_displayname $256
    object_type $100 object_description $256 user_title $200
    parent_id $20 parent_name $100 parent_displayname $256
    parent_type $100 parent_description $256 relationship_type $20
;
set
    gpinfo(in=users)
    gginfo(in=groups)
    idgrps(in=terminating_groups drop=externalkey keyid
        rename=
            (
                objid = object_id
                name = object_name
                displayname = object_displayname
                grpType = object_type
                description = object_description
            )
    );
/* Cleanup */
object_displayname = coalesceC(object_displayname, object_name);
parent_displayname = coalesceC(parent_displayname, parent_name);
if terminating_groups then
    object_type = coalesceC(upcase(object_type), 'GROUP');

/* Define relationships */
if users then relationship_type = 'GROUP-USER';
else if groups then
    relationship_type = CATS(upcase(parent_type), '-', upcase(object_type));
else if terminating_groups then
    relationship_type = 'TERMINATING';
run;

```

The final data set contains the following columns:

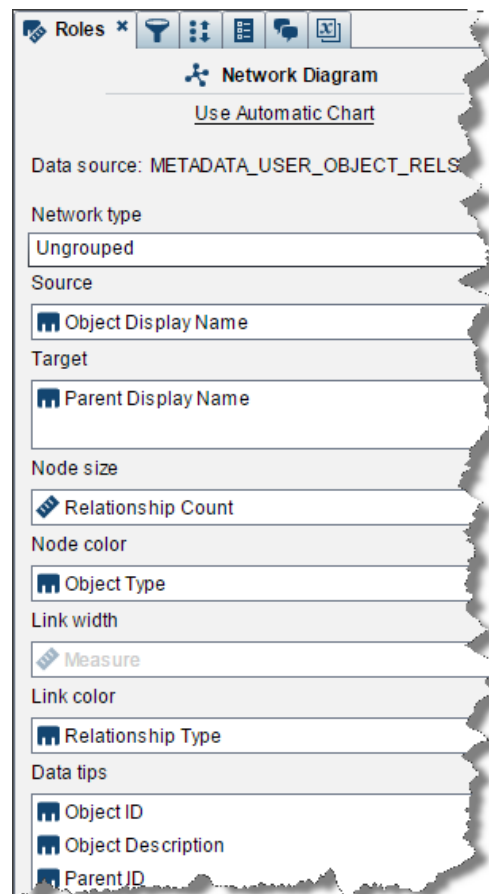
- object\_id – Unique ID from SAS metadata for user, group, or role.
- object\_name – Name of object (user ID, group name, or role name).
- object\_displayname – Display name of user, group, or role.
- object\_type – Type of object (values equal USER, GROUP, or ROLE).
- object\_description – Description provided from SAS metadata.
- user\_title – User title if the object is a user.
- parent\_id – Unique ID of the parent object, either a group or role.
- parent\_name – Name of parent object (if exists).
- parent\_displayname – Display name of parent object (if exists).
- parent\_type – Type of parent object (values equal USER, GROUP, or ROLE).
- parent\_description – Description provided from SAS metadata for parent object.
- relationship\_type – Describes relationship between object and parent (ex: USER-GROUP)
- relationship\_count – Indicator to count relationships.

## ANALYZING USER, GROUP, ROLE RELATIONSHIPS IN SAS VISUAL ANALYTICS

After structuring and cleaning user identity information from SAS metadata, data is loaded into the SAS LASR server for analysis in SAS Visual Analytics Explorer. A network diagram is created with roles defined as shown in Figure 7.

- Network type = Ungrouped
- Source = Object Display Name
- Target = Parent Display Name
- Node size = Relationship Count
- Node color = Object Type
- Link color = Relationship Type
- Data tips = all other columns for reference

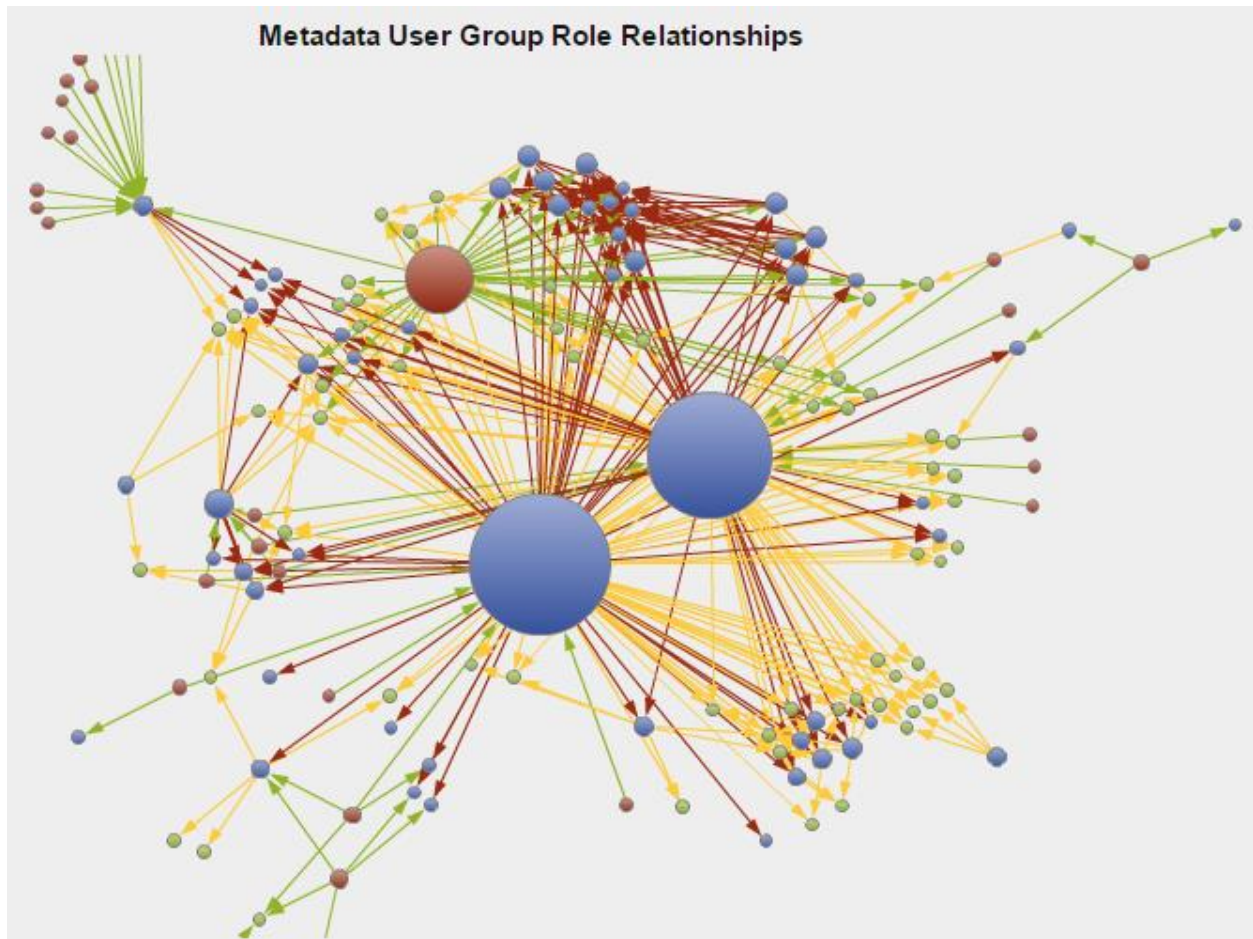
Data tips are useful because they provide reference details when hovering over nodes in the network diagram.



**Figure 7: Metadata User Object Network Diagram Roles.**



The network diagram output is shown below in Figure 8. Blue nodes represent groups, green nodes represent roles, and red nodes represent users. There are clearly two groups, which contain the most relationship connections. The largest red node shows there is a single user, which is a member of the most groups and roles.

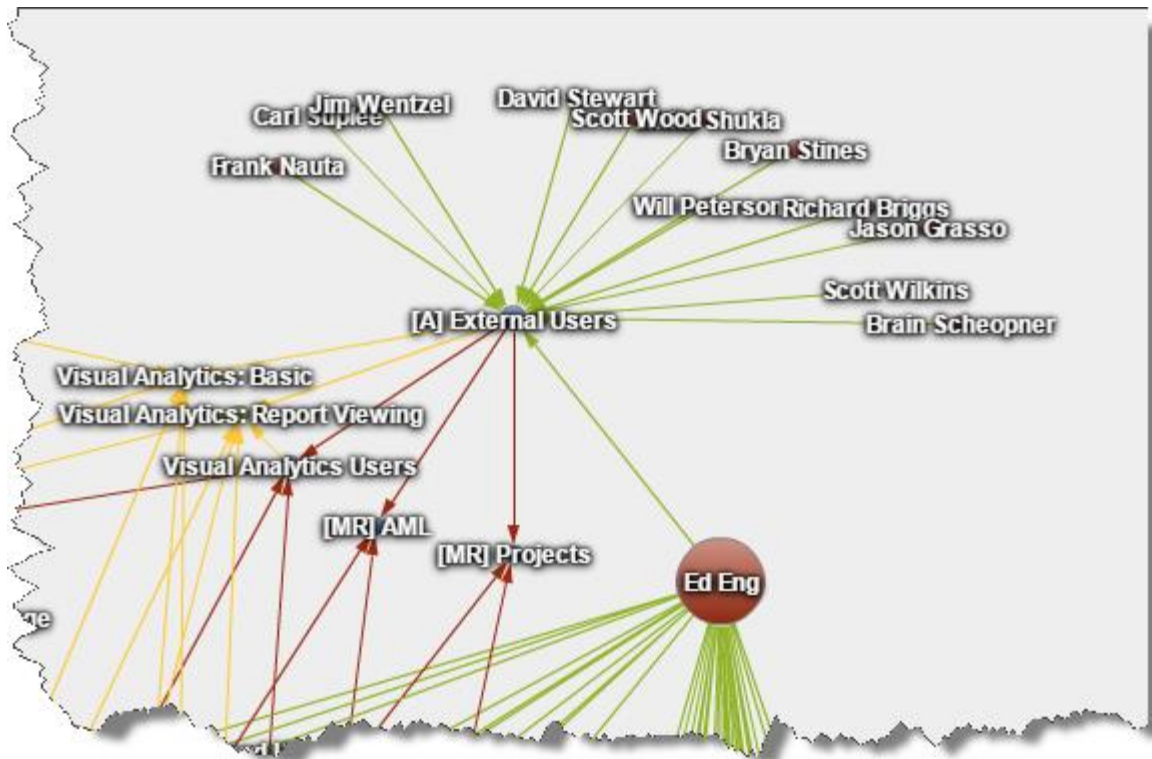


**Figure 8: Visualizing users, groups, and roles in SAS metadata.**

Red lines show group to group relationships. Green lines show user to group relationships. Yellow lines show group to role relationships.

Taking a closer look in Figure 9 below, the network diagram highlights what groups external users inherit and a potential hole in the security model. External users are defined in the [A] External Users group. It appears user Ed Eng is a member of this group but also inherits a number of other groups.





**Figure 9: Highlighting specific group relationships in SAS metadata using a network diagram.**

Visualizing user, group, and role relationships is useful mainly for audit purposes. It can highlight pockets of users that are managed by specific groups as well as accidental inheritance. Network diagrams are also useful because they show the complex nature of some environments due to complex security models.

For additional documentation on the %MDUEXTR macro, refer to the SAS 9.4 Intelligence Platform: Security Administration Guide.

### EXAMPLE: VISUALIZING BANK TRANSACTION TYPE HIERARCHIES

Bank transactions can be categorized into classic hierarchies that describe the nature and type of a transaction. For example, credit and debit can be used to describe transactions at the highest level. Categories such as cash, check, and wire are examples of lower level values within a hierarchy.

#### UNDERSTANDING SOURCE DATA

Bank transactions are very granular and provide specific details such as the customer that performed the transaction, when the transaction occurred, what location the transaction occurred at, the type of transaction, and the amount of the transaction. This demonstration uses a three-level hierarchy to describe the type of transaction:

- 1) **Transaction CDI** – Credit, Debit, or Intermediary. Describes the general direction of the transaction or if it is simply an event on a customer's account.
- 2) **Primary Medium** – Describes the type of transaction such as cash, check, loan payment, or wire.
- 3) **Secondary Medium** – Provides a more specific description of the type of transaction such as the type of check, type of cash, or type of loan payment.

Examples of transaction types used in this demonstration are shown below in Figure 10.

DEBIT	CASH	BASE CURRENCY	\$17,200,474.31
		N/A	\$8,440,354.63
		NON-BASE	\$8,439,489.51
	CASH ADVANCE	N/A	\$8,510,669.42
	CASHED CHECK	N/A	\$8,556,806.14
	CD	N/A	\$8,626,388.76
	CHECK	BALANCE TRANSFER	\$8,474,789.27
		BUSINESS	\$8,428,728.75
		CONVENIENCE	\$8,484,730.42
		COUNTER	\$8,449,451.44
		N/A	\$13,821,676.13
		PERSONAL	\$8,581,713.71
		SAVINGS BOND	\$17,241,118.35
	FEE	ACCOUNT	\$8,628,444.58
		FINANCE CHARGE	\$8,615,060.86
		RETURNED CHECK	\$8,540,530.78
	TRANSACTION		\$8,527,777.77

Figure 10: Example transaction type hierarchy.

## BUILDING A HIERARCHICAL NETWORK DIAGRAM OF TRANSACTION TYPES

After structuring and cleansing source data transactions, data is loaded into the SAS LASR server for analysis in SAS Visual Analytics Explorer. A network diagram is created with roles defined as shown below in Figure 11.

**Network type**

Hierarchical

**Levels**

- Transaction CDI
- Primary Medium
- Secondary Medium

**Node size**

Measure

**Node color**

Any data item

**Link width**

Amount

**Link color**

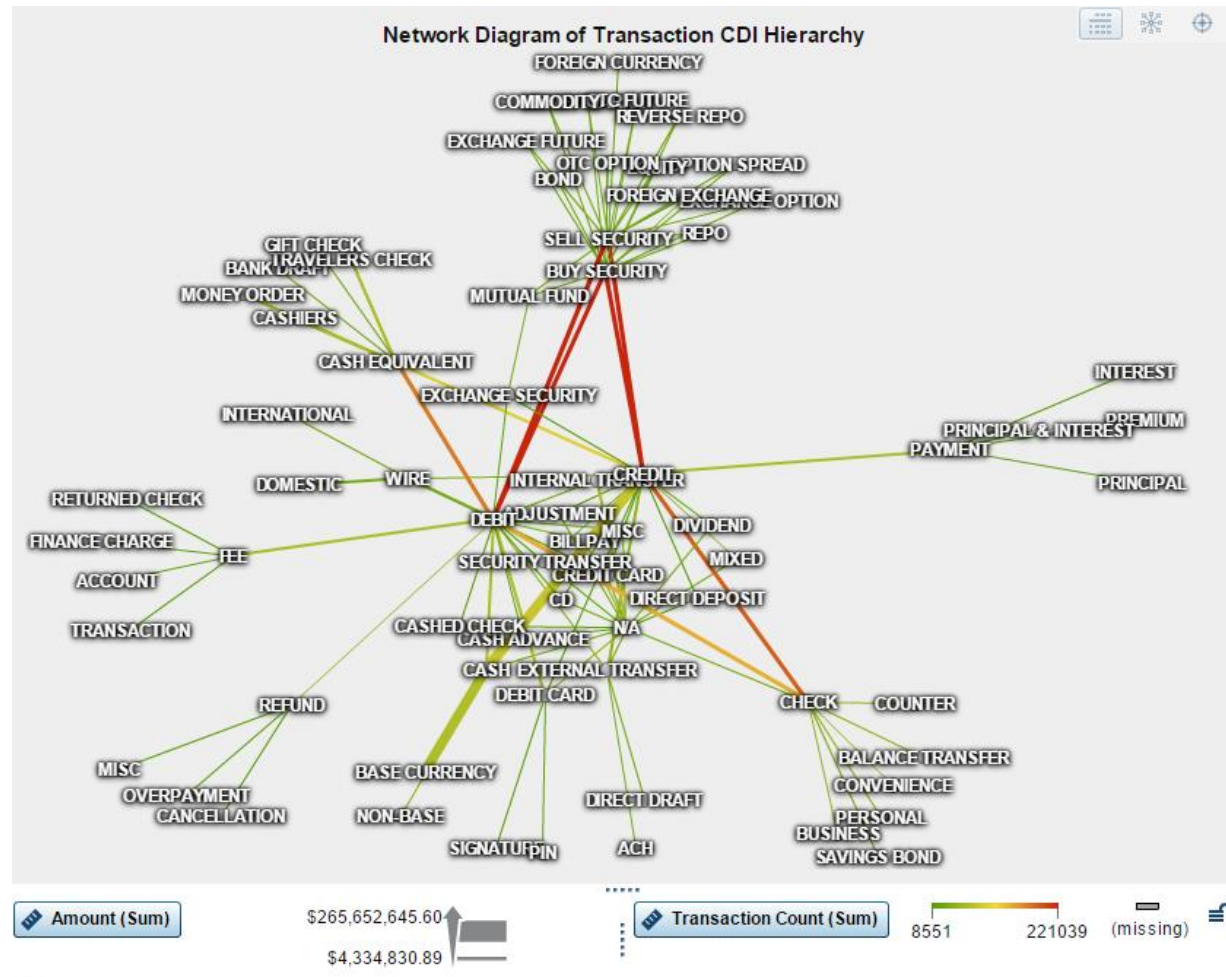
Transaction Count

**Data tips**

Any data item

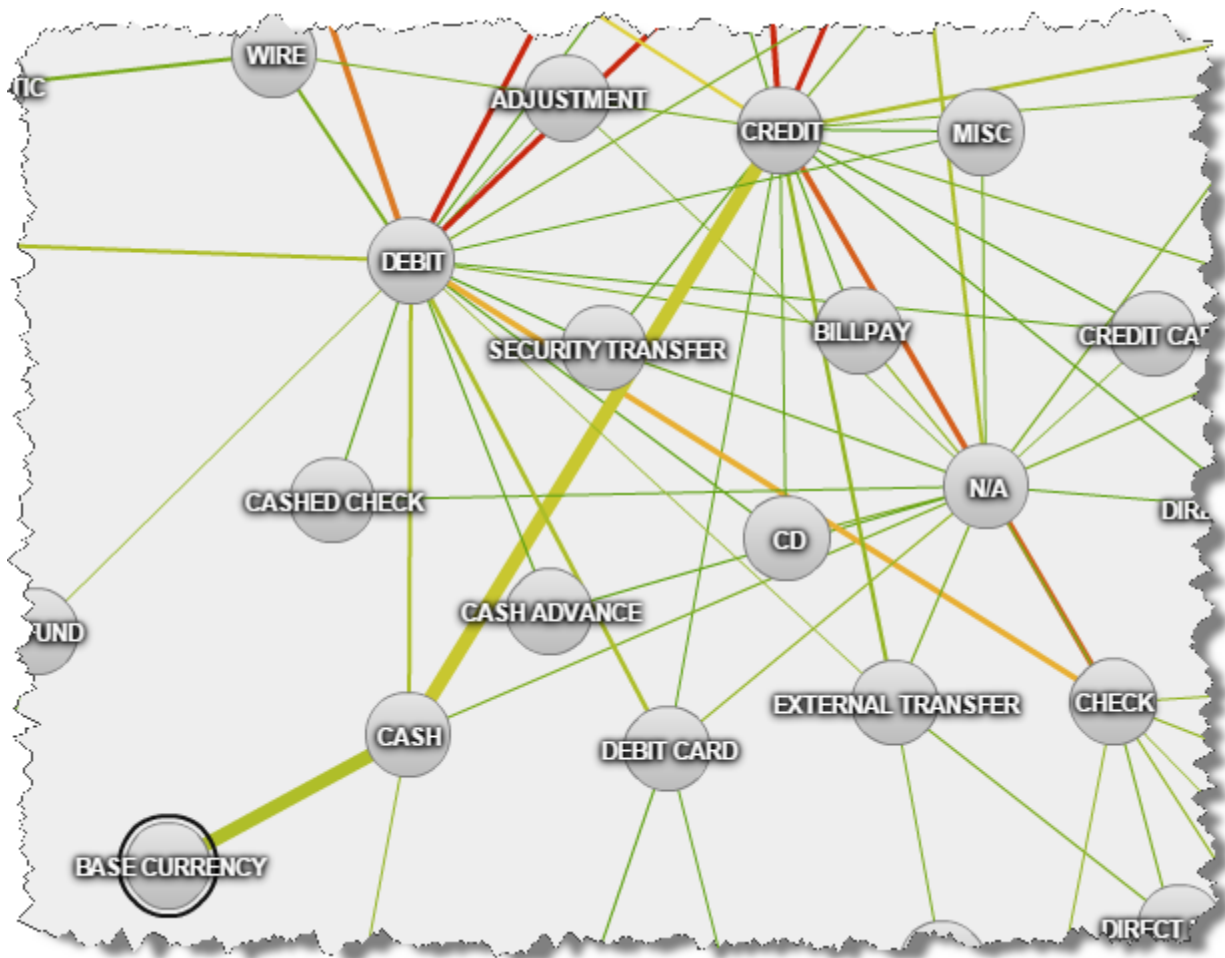
Figure 11: Network diagram roles for transaction type analysis.

The network diagram output is shown below in Figure 12. The thickness of the line represents the sum of transaction amount while color represents the transaction volume.



**Figure 12: Transaction type analysis using a network diagram in SAS Visual Analytics Explorer.**

At a high level, this network diagram shows us that the highest volume of transactions occur within buying and selling securities, or investments. This may be helpful in assessing a bank's risk or exposure for certain types of transactions. The highest magnitude of transaction dollar amounts occurs with cash as shown below in Figure 13. This provides further context to understand the behavior of customers and how money is spent.



**Figure 13: Closer look at the highest volume of cash transactions.**

Drilling into particular nodes of interest allows additional analysis to be performed using other visualizations; such as a table listing, which would be a simple approach to visualize specific transactions at a more granular level. Right clicking on a node (or multiple nodes) and selecting “Include Only Selection”, filters the visualization by adding the selections to the filter criteria. Once node values are filtered, the visualization can be switched to dig deeper into the selected node values.

### EXAMPLE: VISUALIZING FOREIGN BANK WIRES

In the banking industry, specifically within the Anti-Money Laundering (AML) space, it is important for financial institutions to understand exposure and risk for sending and receiving foreign wires to and from other countries. SAS Visual Analytics Explorer can be very useful in starting an analysis with network diagrams to accomplish this. In contrast with the previous two examples, the nature of source data in this demonstration can be more of a hybrid approach.

### UNDERSTANDING SOURCE DATA

The nature of bank wire transactions lends itself nicely to either ungrouped or hierarchical network diagrams. Bank wire transactions contain attributes which describe the originator of the transaction, the beneficiary of the transaction, and sometimes an intermediary for the transaction. The originator is the bank which sends the funds, the beneficiary is the bank which receives the funds, and the intermediary can sometimes be a bank which receives funds on behalf of the beneficiary bank. Ungrouped diagrams technically work because the complexity of data is very simple and there are “source” and “target” variables. This demonstration utilizes a hierarchical approach to demonstrate the use of filtering on

nodes, which is more straightforward using a hierarchical approach. This example also assumes additional AML attributes are available such as risk scores for the transaction and country risk rating.

To begin the analysis, source data is gathered and validated. The grain of the data is at the transaction level, which is the lowest level possible (also referred to as the atomic level). Key variables such as country, risk score, and dollar amounts are present.

The final data provides comprehensive detail on columns used in this example:

- Beneficiary country – where the foreign wire was sent to, i.e. who “benefited” from the wire exchange.
- Originator country – where the foreign wire was sent from, i.e. who sent the funds.
- Alerted indicator – a simple indicator to show that the transaction triggered an AML alert.
- Country Risk – discrete risk category for the country where origination occurred.
- Amount – exact dollar amount of the foreign wire.
- Score – numeric value of the risk classification determined by an AML solution. Note, this is specific to the solution used at an individual bank.

Example data is shown in Figure 14 below.

Originator Country ▲	Beneficiary Country	Amount	Alerted Ind	Beneficiary Country Risk	Risk Score
AFGHANISTAN	MEXICO	\$233,316	0	HIGH	8
AFGHANISTAN	SINGAPORE	\$903,775	0	HIGH	8
AFGHANISTAN	UNITED STATES	\$21,596	1	HIGH	8
AFGHANISTAN	UNITED STATES	\$400,132	0	HIGH	8
AFGHANISTAN	SINGAPORE	\$725,512	0	HIGH	8
AFGHANISTAN	UNITED STATES	\$10,545	0	HIGH	8
AFGHANISTAN	UNITED STATES	\$474,250	0	HIGH	8
AFGHANISTAN	CANADA	\$618,622	0	HIGH	8
AFGHANISTAN	UNITED STATES	\$387,221	0	HIGH	8
AFGHANISTAN	BRAZIL	\$548,226	1	HIGH	8

Figure 14: Foreign bank wire data for network diagram analysis.

## VISUALIZING FOREIGN BANK WIRES IN SAS VISUAL ANALYTICS EXPLORER

Data is loaded into the SAS LASR server for analysis in SAS Visual Analytics Explorer. A network diagram is created with the following roles:

- Network type = Hierarchical
- Levels = Beneficiary Country and Originator Country
- Node Size = Count of alerted indicator (measure)
- Node Color = Country Risk
- Link Width = Amount
- Link Color = Median score

The network diagram output is shown below in Figure 15. Larger nodes indicate higher volume of alerted transactions originating from a specific country. Lines that are redder indicate a higher threshold of risk scores for transactions flowing from the originating to beneficiary country.



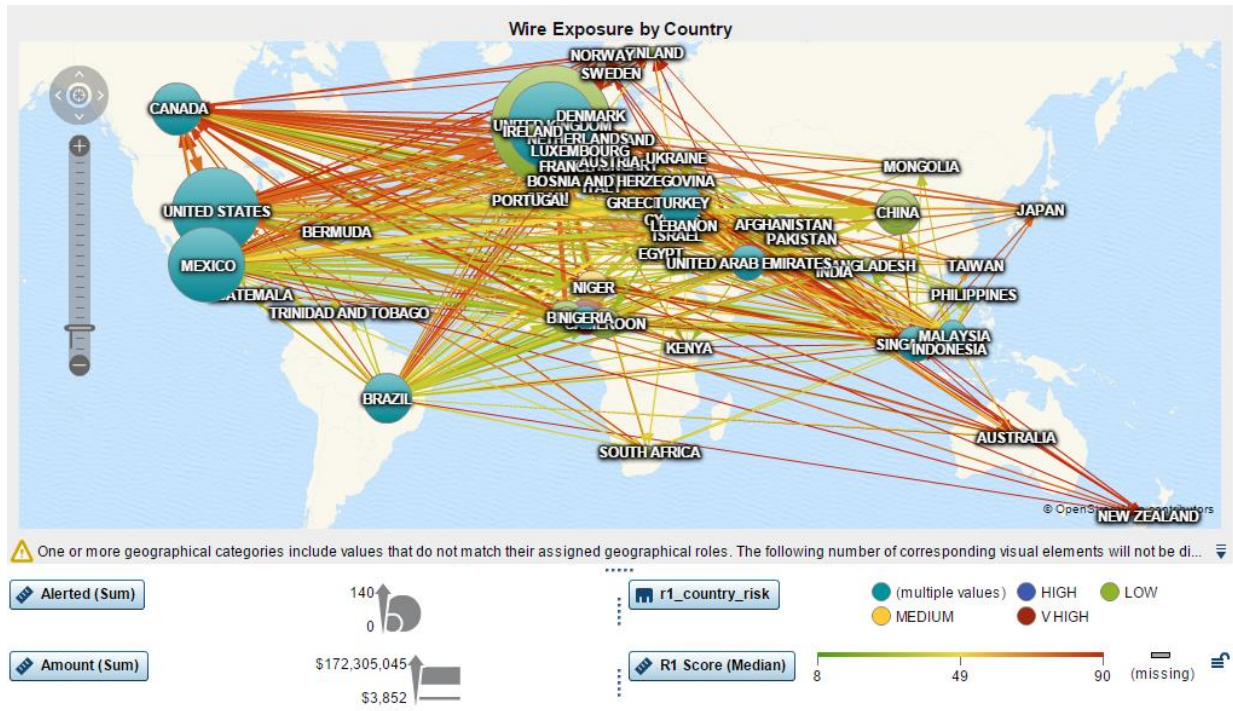


Figure 15: Foreign wire exposure by country at a global level.

At a high level, this example shows the power of being able to visualize an entire population of values as they relate to one another. Rather than having to make a guess at which country to focus on, a network diagram allows the analyst to visualize by using all data available first, then make appropriate decisions based on an overall perspective. One method to dig deeper into network diagram data is to right click a node (or nodes) and select "Include Only Selection". This automatically selects values for the respective nodes and includes in a visualization filter. This can be confirmed under the Filter tab of the exploration. Figure 16 shows an example of isolating the country of Ukraine using the "Include Only Selection" option.

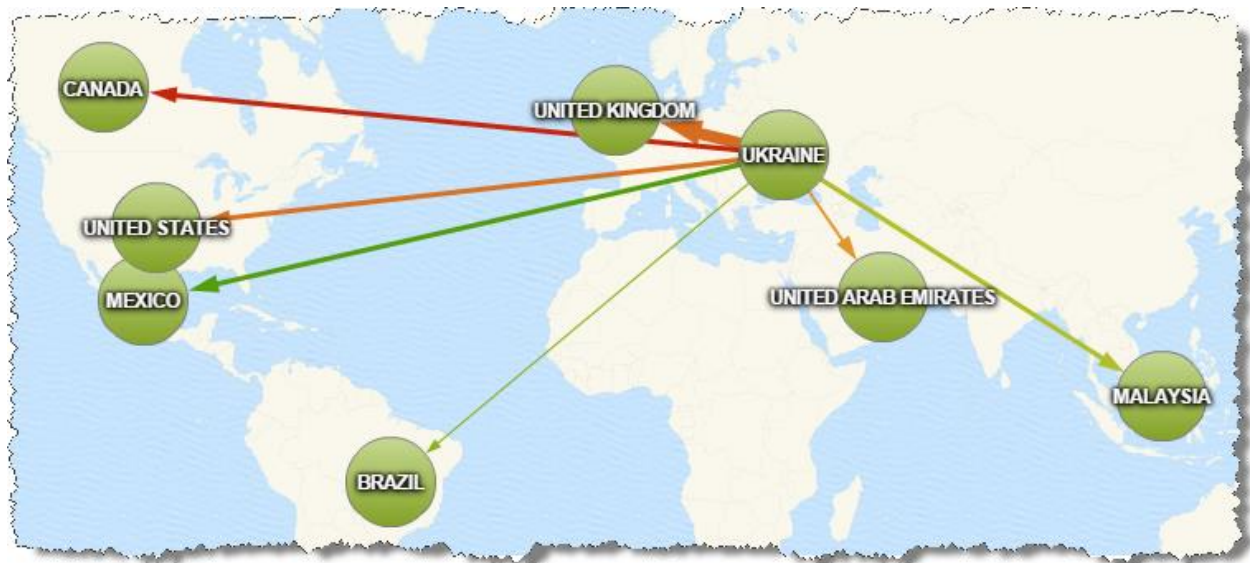


Figure 16: Isolating a specific country for wire exposure analysis.

Understanding the exposure of foreign wires is important for a bank and a common task handed down by regulators. From here an analyst may dig further by switching the network diagram to another



visualization for further analysis. A heatmap or correlation matrix may provide additional analysis of variables within the data; or a table listing would provide the specific transaction detail. Due to how the network diagram functionality works, the “Include Only Selection” option only passes filter values to the local visualization and not the entire exploration. Therefore the same visualization has to be reused to drill into the data of a network diagram.

## CONCLUSION

SAS Visual Analytics Explorer makes it very easy to create network diagrams, assuming the source data is structured properly. A network diagram can be a great tool for finding the needle in the haystack. Conceptually speaking, rather than picking a subset of data to analyze, consider analyzing all available data to start using a network diagram if the source data permits, allowing you to quickly identify relationships in the data that can be used for further analysis through a simple click.

## RECOMMENDED READING

- *SAS® Visual Analytics User's Guide*
- *SAS® 9.4 Intelligence Platform: Security Administration Guide*

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Stephen Overton  
Director of Financial Solutions  
Zencos Consulting  
(919) 341-9667  
[soverton@zencos.com](mailto:soverton@zencos.com)  
<http://www.linkedin.com/in/overton>  
<http://www.zencos.com/>

Ben Zenick  
COO  
Zencos Consulting  
(919) 459-4603  
[bzenick@zencos.com](mailto:bzenick@zencos.com)  
<http://www.zencos.com/>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.