4th ICAAAE 2013 SEPTEMBER 22-25, Hammamet Tunisia
4th International Conference of the African Association of Agricultural Economists
icaaae.org
aaae-africa.org

The use of social network theory in value chain analysis

David Guillemois
david.guillemois@gmail.com

Invited panel session sponsored by the International Livestock Research Institute (ILRI) through the CGIAR Research Program on Policies, Institutions, and Markets (PIM)

"Network Analysis Applied to Livestock Value Chains: Relationships beyond Demand and Supply and Their Contribution to the Impact of Upgrading Interventions"

The use of social network theory in value chain analysis

David Guillemois[1]

Introduction

"*Social network analysis is a research methodology that focuses on relationship between and among entities, and on the pattern and implications on these relationships*"[i]. Social network theory is becoming more popular in data analysis as it enables to develop a more detailed understanding on the causality effect through analyzing how entities (being individuals, subject, organization, etc.) relate to each other. This capacity to understand and calculate network related data is of particular interest for research focusing on value chain as the chain is a network. Understanding how entities composing this network are able to influence the value along the chain contributes to value chain analysis.

Research is about isolating key determinants linking specific causes to specific effects. Ability to isolate or to guess some hypothesis of causality will orientate the research toward key determinant. Using a different angle of data presentation contributes to better inform the researchers on the directions of further data analysis. The display of data related to value chain as visual network and the calculation related to social network provide complement the classic database analysis.

Social network analysis theory

As it was defined by Nooy *et al*[ii], social networks analysis focuses on ties among, for example, people, groups of people, organizations, and countries. According to some authors, the individual is not the basic social unit, but social atoms consisting of an individual and his or her social, economic, or cultural ties. The ties can represent friendship, buyer-seller relations, contracts, certifications, cooperation, and so on[iii] and these are some of the actors composing the value chain.

The main goal of social network analysis is detecting and interpreting patterns of social ties among actors[iv]. How actors interact together along the value chain will then be interpreted on the basis of who are the actors and how are they connected.

A network is based on four components:

- The entities composing the network (also called actors, vertices, nodes)
- The link between these entities and their characteristics (also called edges)
- The profile of each entity (this information can be a category, a vector, etc.)
- The characteristic of the link between each entity (this information can be a category, a vector, etc.)

These four levels of information are sufficient to analysis the network and to create a visual network. Filters on specific information foreseen to be displayed, will structure thematic networks (focus on region, on categories of edges, on categories of actors, etc)

Some key points of social network theory are:

- **The closeness** is about measurement of links separating specific actors. How many direct or indirect links are needed to connect one actor to another. A central actor (highly connected) will be closer to most actors than an actor on the outskirt of the network (lower level of connections).
- **The degree** is the number of connections hold by one actor, contributing to increase its centrality status within the network. The actor with the most connections has as the highest degree in the network. The number of connections will depend on the type of connections we will be looking at. So, while connections

[1] Consultant. Email : david.guillemois@gmail.com

have a binary status for the degree measurement, their respective value (type of connections, profile of actors connected) will be appreciated during power and influence measurement.

- **The betweenness** centrality quantifies the number of times an actor acts as a bridge along the shortest path between two other actors. So, actors with high level of betweenness are key influential actors and "entry points" or key facilitators linking different groups of actors as one single network.
- **The centrality** is the degree to which an actor occupies position close to the center of the network. The centrality is actually based on the number of links with other actors and the closeness to all other actors.
- **Homophily** is the tendency to relate to actor with similar characteristics. A network where homophily is important will depend on few actors with high level of betweenness to link thematic sub network together.
- **A network density** is the ratio of the number of edges in the network over the total number of possible all pairs of nodes.
- **A node's clustering coefficient** is the density of its neighborhood. The clustering coefficient for an entire network is the average of all coefficients for its nodes. Clustering is indicative of the presence of different characteristics of the sub groups.
- **The influence** of actors composing the network is closely related to centrality status of actors but also considers qualification of the links and the level of influence of connected actors. Influence is therefore also about being closely and strongly connected to others influential actors.

Application of SNA to dairy value chain analysis

Social network analysis has been used during a dairy value chain survey in Uganda and Tanzania along a classic survey. While the sampling followed the classic statistical approach, the questions have been structured to code each actor as an unique entity in the database.

Approach used for data coding (to link entities)

Field data collection has focused on questionnaires with milk traders, milk producers and business development services providers (BDS). During each interview, respondents have been asked about the specific names of their customers/buyers, specific services provided/received and the status of their relationship with these actors (quantity of milk sold/bought, mode of payment, type of training provided/received, etc). All names mentioned have then been coded to create the list of actors composing the global network and links between actors have been added to the database.

Approach used for sampling during field survey and its limits

The sampling of entities interviewed is critical to appreciate what the network would be. A classic survey focus on a sample group representative of the whole group assessed. The representativity of this sample is mainly aiming at using its proportional responses foreseen to be valid for the whole group.

- The network analysis theory displays the data as a whole and does not focus on proportional responses. Therefore, the use of the network analysis theory for the data collected during field survey using a representative sample should be appreciated within these specific limits. The network structure of a "sample" does not represent the network of the whole group and therefore the network is broken and not as dense and connected as it is in the reality. The traditional sampling of the representativity of the group assessed is not appropriated to use the network analysis theory to its maximal capacity. However, some key information within the sample group, and therefore within understood and accepted limits, can be drawn. A field data collection for a value chain analysis cannot interview all actors. Therefore, the use of network analysis theory in value chain analysis will always be working with limited sample and the network will remain "broken" compared to the reality. However, some network analysis can be done within sub networks.

- A classic survey on the field shows average, percentage, sum, graph which should be understood as vertical calculation within a database where each questionnaire is a line and each question is a column. The network analysis is rather looking at connections at each actor level and therefore provides a rather horizontal analysis of the same database. Coupling these two approaches enables to get richer understanding of the data analysis.
- Applying the social network theory to the dairy value chain provides information on the social dynamism along the dairy chain. This enables to better understand how the chain is working, who are the key entities influencing the chain, what are the best entry points to generate cascading effects along the chain, who will mainly benefit from changes (risk of monopoly of markets, isolated actors (low centrality), etc). Network analysis can further inform if some entities could be reluctant to change especially they occupy a dominant position in term of betweenness.

Software to analyze the network

Different software using network analysis can display the network and can do calculation related to network analysis. For instance we can cite:

- Netminer[2]
- UCINET[3]
- Pajek
- Gephi[4]
- Sentinel visualizer[5]
- Netmap[6]
- MS Excel Node XL[7]. it is an adds-in to MS Excel

For the analysis of the network, we chose Pajek software (*Program for Large Network Analysis*) mainly due to the simplicity to code the data to be displayed than for others reasons. It would be fair to mention that a detailed knowledge of all other social analysis software has not been conducted. Learning how to use these softwares and understanding the details specifications is particularly time consuming.

Pajek is well known within the social analysis open-source software for its capacity to handle large volume of data and was then foreseen to be a fair choice. Most of the effort has then been to prepare the databases as a file compatible with Pajek.

Interface

The database of information collected for this dairy value chain in Uganda and Tanzania can be displayed as a whole network through Pajek software but has been redesigned to present the data set per field of interest (per country, per region, per category of links, per profile, etc). These filters have enabled to develop a set of thematic networks. An interface between the database and Pajek file was then needed to recode automatically the entities and their links prior importation of the data sets into Pajek.

The data collected during the survey has been entered in Microsoft Access as a set of tables and classic analysis has been handled by software (SPSS (Statistics Package for Social Scientists) or Microsoft Excel). All data has then been imported under MS Excel to create columns with information on the link between vertices.

The interface has been focusing on:

[2] http://www.netminer.com/index.php
[3] https://sites.google.com/site/ucinetsoftware/home
[4] https://gephi.org/
[5] http://www.fmsasg.com/
[6] http://info.iet.unipi.it/~luigi/netmap/
[7] http://nodexl.codeplex.com/

1. Consolidation of all key information with a focus on data related to entities profile (identification of actors, qualification of actors, vector information per actor) and qualification of connections (identification of links, qualification of links, vector information per link). The database has been structured to get one unique actor per line and all its connections on the same line.
2. Preparation of filter for the data to be used in the network. Filters were foreseen to be based on the region of data collection (Uganda or Tanzania), category of respondents (producers, traders, services providers) but can be related to others type of analysis (focus on milk trading, services provision, training). The interface is also seeking information in an interactive manner in the various tables composing the database. This induces that updating of the network databases through data standardization or setting of score or new category per link or actors can be done automatically.
3. Standardizing the coding of all actors as per Pajek structure, after the application of specific filters.
4. Standardizing the coding of the links between actors, based on Pajek coding. Connections of interest were notably related to milk exchange, training services, mode of payment, etc.
5. Standardizing specific vector information for the qualification of entities but also for their links (vector files, quantity of milk produced, quantity of milk sold/procured, etc.)

The key structure of Pajek file is to code all entities using incrementing approach and then to define each link through an information linking two actors as a link between their attributed numbers.

Table1. *Example of the pajek file coded.*

The main PAJEK file indicates the number of vertices and then provides a unique number from 1 to 280 (for this example) to each vertice. Shape of each vertices, border, and color can be indicated for each line. Within the same file, the link between vertices is indicated under "edges" category. The vector information on the quantity of milk exchanged and specific category of line and color are then mentioned here.	File vector. Information indicates a vector information for each vertice. These vectors information have been used to quote the quantity of milk per producers or traders.
**vertices 280* *1 "PU100" ellipse ic Red* *2 "PU101" ellipse ic Red* *3 "PU102" ellipse ic Red* *4 "PU103" ellipse ic Red* *5 "PU104" ellipse ic Red* *6 "PUG09" ellipse ic Red* *7 "PUG103" ellipse ic Red* *8 "PUG104" ellipse ic Red* *9 "PUG105" ellipse ic Red* *10 "PUG106" ellipse ic Red* *…. / ….* **edges* *201 50 37 w 36 c Black* *201 149 127 w 126 c Black* *202 165 37 w 36 c Black* *202 166 27.5 w 26.5 c Black* *203 120 42 w 41 c Black* *203 147 32 w 31 c Black* *204 34 28.5 w 27.5 c Black* *204 144 32 w 31 c Black* *205 25 27 w 26 c Black* *… / …*	**vertices 280* *1 63* *2 215* *3 23* *4 97* *5 53* *6 82* *7 1* *8 1* *9 1* *10 1* *… / …*

Remarks

The filtering of the database to extract very specific information of the network do generate some "orphans" within the network i.e. entities not excluded by the filtering system but exclusively connected to actors excluded by the application of filters. The example focusing on milk flow excludes the BDS, so all producers or traders reported to be exclusively connected to BDS will appear without any connection.

The field data collection has been based on connecting entities by mentioning all of them during specific interviews with respondents. So, some interviews with respondents have mentioned others entities which have not all been interviewed. This point is important as we do have some entities with less information. The entities "mentioned" but "not interviewed" are not supported by detailed information on their respective sets of connections, qualification of connections, and details on their profile. This aspect of different level of information available per actors is critical in interpreting the network especially on its respective connections and on all vectors related information. On some networks, it was useful to mention which actors have been interviewed versus which ones have only been "mentioned as a connection" during interviewed using a different color code.

The display of the network is never the same. Refreshing the calculation change how the computer display the network. Key characteristics will remain the same but appearance can evolve.

Preliminary results

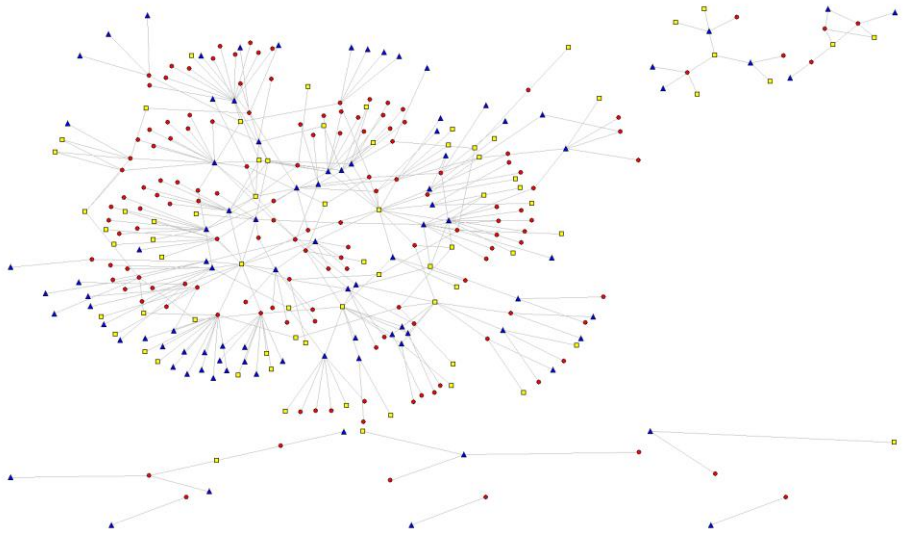

Figure 1. *Comprehensive network for Mwanza-Tanzania showing producers (red circles), traders (blue triangles), BDS (yellow squares) and all their respective connections.*

Density [loops allowed] = 0.01072156. *Calculated by Pajek.* The loop allowed density is the number of loops (path between vertices enabling to link back the first vertice of this path) divided by the number of possible loop. So, the more connected are the BDS, the higher would be the density of the loop in the network.

Average Degree = 3.36977492**.** *Calculated by Pajek.* As an average each vertice has 3.37 connections.

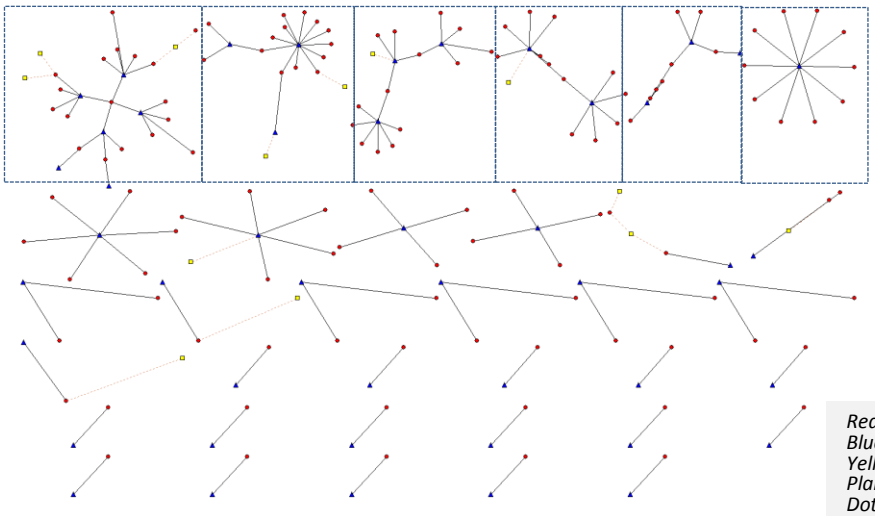

Figure2. *Mwanza connections between traders, producers and BDS. Connection are related to milk exchange and training*

As mentioned earlier, the approach to "sampling" used during field data collection generates a fairly broken display of network but some conclusion can still be made from the cluster of network showing specific characteristic of connections.

1. There is no connection trader-trader.
2. Very few producers are selling to more than one trader putting the producers in a fairly weak position the value chain.
3. BDS in Mwanza do not have a central role and in most of the cases provide support to the most isolated entities (low level of connections and on the outskirt of the network) rather than to the central and highly connected entities. Some considerations, on impact of training along the chain could then be considered.

Connectness per actor is calculated based on the possible way to reach any actor starting from any actor by traversing edges. The network must therefore not be broken. Connectness can then be calculated only for smaller but unbroken network i.e. a connected component where all actors are lined directly or indirectly to each other.

Network **betweenness centralization** = 0.01406224 for the graph above. Given the type of actors in the network and the flow of good along the chain, the BDS are foreseen to be the main actors increasing the between centralization of the network.

Degree

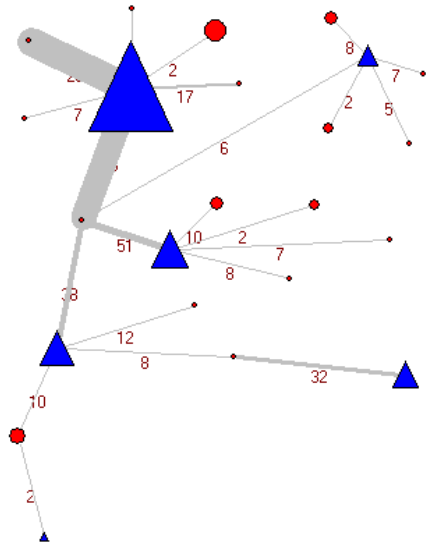

Figure3. *Snapshot on part of the network milk production selling in Mwanza, Tanzania.*
- *Blue triangle represents milk traders.*
- *Red circle represents milk producers.*
- *Size of the triangle and circle represents respectively the volume of milk procured and produced.*
- *The size of the link represents the flow of milk selling.*

Focus on one section of the network on the milk selling in Mwanza, Tanzania provides key information; with exception to one milk producer foreseen to be a major producer (given the level of milk exchanged) all producers within this network are connected to only one trader. The milk producers selling to more than one trader also sell to the larger trader within this cluster.

The largest trader of this network is buying milk from two large producers.

For the value chain analysis it would be interesting to better appreciate the connections between traders to understand the transaction of milk between small and large traders.

Betweenness

The network representing milk exchange and training connections in Mwanza-Tanzania between milk producers, milk traders and all connections with the BDS (Figure 1) compared to network where the BDS are excluded shows the capacity of the BDS to connect the network. This finding is not a surprise as the BDS need a larger number of customers, and can propose a large sets of different services. So, the BDS are the entities with the highest level of betweenness as the graph indicated as figure 1 below shows much higher level of density of the network when BDS are fully included.

Conclusion

The social network analysis theory can be applied to research field but with limitations as the field approach to sampling of the group assessed breaks the network and not all actors can be met for an interview. However, trend identified even within a broken network can be used as a direction of research to be confirmed with classic data analysis.

Social network analysis can be used to get a different perspective in data analysis. This type of analysis is particularly important in value chain analysis as the chain is a network.

Visual representation of broken network might be within the most broken network, the main outputs. “The human eye is trained on pattern recognition”v and the visual representation of the thematic network do provide a detailed data analysis.

The approach to sampling of the field data collection could be understood differently with interest to concentrate data collection in specific areas (geographic, sectorial, etc.) to get more data within specific

networks increasing network density and therefore improving accuracy of network representation and interpretation. Less dispersed sample would enable to focus data collection and will enable to better use the potential of calculation of the social network analysis theory.

References

[i] "Introducing network analysis as a research technique" by Adjunct associate professor Graham Durant-law CSC, Phd. University of Canberra.

[ii] Nooy et al. (2005)

[iii] Wasserman and Faust, 1994; Nooy et al., 2005

[iv] Exploratory social network analysis with pajek. Wouter De Nooy, Andrej Mrvar and Vladimir Batagelj.

[v] Exploratory social network analysis with pajek. Wouter De Nooy, Andrej Mrvar and Vladimir Batagelj.