DOI: 10.54597/mate.0065
Horváthné Kovács, B., Pintér, Zs., Nagy, M. Z. (2022): Network analysis solutions in the agri-food sector. In: Srečec, S., Csonka A., Koponicsné Györke, D., Nagy M. Z. (Eds.): Management of agri-food chains. Gödöllő: MATE Press, 2022. pp. 96–110. (ISBN 978-963-623-023-4)

CHAPTER 7

Network analysis solutions in the agri-food sector

Authors:
Horváthné Kovács, Bernadett ORCID: 0000-0002-2038-6428, Hungarian University of Agriculture and Life Sciences
Pintér, Zsófia ORCID: 0000-0001-5250-2115, Hungarian University of Agriculture and Life Sciences
Nagy, Mónika Zita ORCID: 0000-0003-0847-190X, Hungarian University of Agriculture and Life Sciences

7.1 Definition, concept, mainstream applications

Deriving from social network (analysis) network studies and science have evolved to many areas of life, such as DNA mapping, logistics or marketing researches.

When thinking about the world in terms of different overlaying networks that connect and transfer friendships, information, money, and power – it becomes obvious that looking at things through the analysis of social networks can lead to a new realization on many interesting topics. Just think some of the following examples that are commonly understood as network concepts, such as online or more traditional social networks of people, degrees of Kevin Bacon the actor, or the way how Facebook algorithms predict products or friends to be offered. We have an intuitive sense that the connections of the people around us are a huge factor[1].

Why to rely on network analysis?
The standard statistical methods would not be effective enough without looking at connections of the social network[2]. Similarities and differences between isolated data points do not, but social networking data analysis gives us tools to quantify those connections between individual points so that we can find patterns in the forces that connect us as a society. If the researchers can find out how one person connected to or disconnected from people, groups, and trends in a population and all those people who seems to be friends with everyone, they are able to reveal individuals in populations that bridge social groups.

In a more practical way of looking at the topic, it is interesting for specific decision makers to gain information on what makes a group of strangers start to form statement groups, what networks are firm, how things like power, beliefs, or even an outbreak of disease flows through the individual connections. These practical questions may be atrgete with quantitive answers and new insights with social network analysis.

How network analysis and science evolved?

Social network analysis is a very open field and there are lots of technical options to try out. Like adding geographic mapping data to understand how physical environments change network dynamics. For businesses on online social media, understanding how people connect (react or otherwise learn) your busniess activities or what information on these people is available can be crucial in your business prosperity.

The rise in computation and emergence of mass of new data sources facilitated social network analysis. Social network analysis is the application of network theory to the modeling and analysis of social systems. It combines the tools for analyzing social relations and the theory for explaining the structures that emerge from these social interactions.

Social networks are studies as ones composed of individuals and organisations, and the aim of the analsis is to quantitatively describe these entities in a formal mathematical language of statistical analysis. Network science adds information on cause-and-reason analysis by capturing the most important feature of social reality that is the relations between individuals.

Network science analyzes empirical data and develops theories to explain the patterns observed in these networks. Such questions are asked as the degree of connectivity within a network, its overall structure how far something will diffuse or propagate through it, or the influence of a given node within the network[3].

7.2 Network analysis in agrifood sector

Applications of network analysis

Social network analysis has already been used to study the *structure of influence within corporations.* Findings might be surprising when modeling the actual flow of information and communications as a network gives a very different picture on seemingly irrelevant employees within the hierarchy who can in fact have significant influence within the network.

Researchers also study *innovation as a process of diffusion of new ideas across networks* where the overall structure to the network is degree of connectivity, centralization or decentralization.

Network dynamics that is how networks evolve over time is another important area of research for example. Social network analysis is used to study the change in structure of terrorist groups to identify the changing relations through which they are created, strengthened or dissolved.

Social network analysis has also been used to study the patterns of segregation and clustering within international politics and culture. By mapping out the beliefs and values of countries and cultures as networks we can identify where opinions and beliefs overlap or conflict. This can be useful for international companies outreaching to sveral cultures with their subdivisions.

Social network analysis is a powerful new method that allows us to convert often large and dense datasets into engaging visualizations that can quickly and effectively communicate the underlying dynamics within the system.

Social network analysis is offering a huge potential for a deeper, richer and more accurate understanding of the complex social systems that make up our world[2].

7.2.1 Networks of the socio-environmental-economic production space

First of all we need to understand how networks are complied. In order to see clearly the elements of a network this chapter shortly introduces the basic vocabulary and statistics of network structures.

Nodes

Nodes or points in a network or diagram are elements at which lines or pathways intersect or branch. In the below articular case there are two nodes (Node2 and Node5) linked.

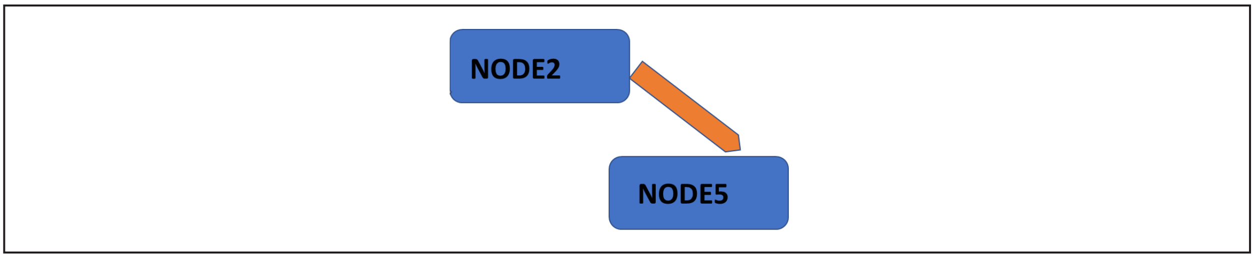

Figure 1. *Pair of nodes with link*

Edges

Edges or links are the pathways between or across the nodes of a network. For instance, in the below graph, Edge1 goes from Node1 to Node2 and so on.

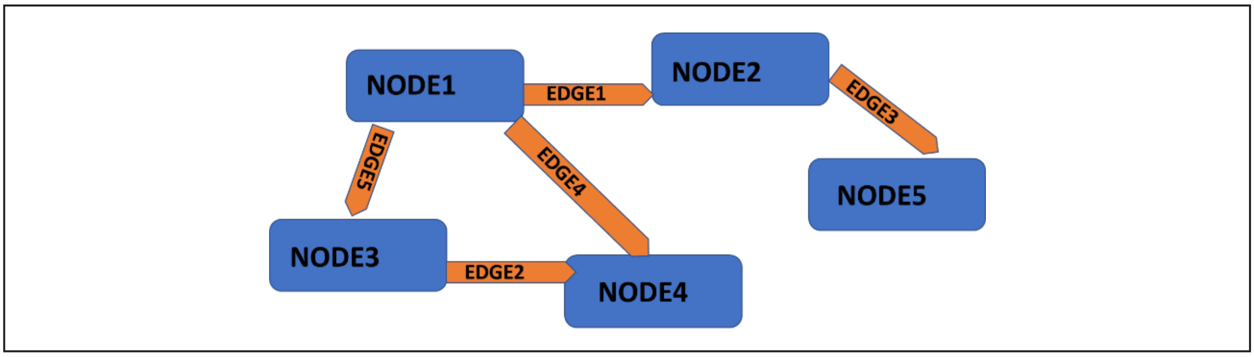

Figure 2. *Links of nodes in a simplified model*

Degrees

The term degree expresses the number of links or edges reaching the intersects presented by the nodes; it measures the direct connections of a node with other ones. Actually the degree of a node gives an insight how well this particular node is connected to the others. In the above example, Node1 has 3 degrees as it is connected directly to three further nodes (Node2, Node3 and Node4).

Degree distribution

A common way that we describe networks is by giving a degree distribution. The degree distribution is simply a tally of how many nodes have each degree.

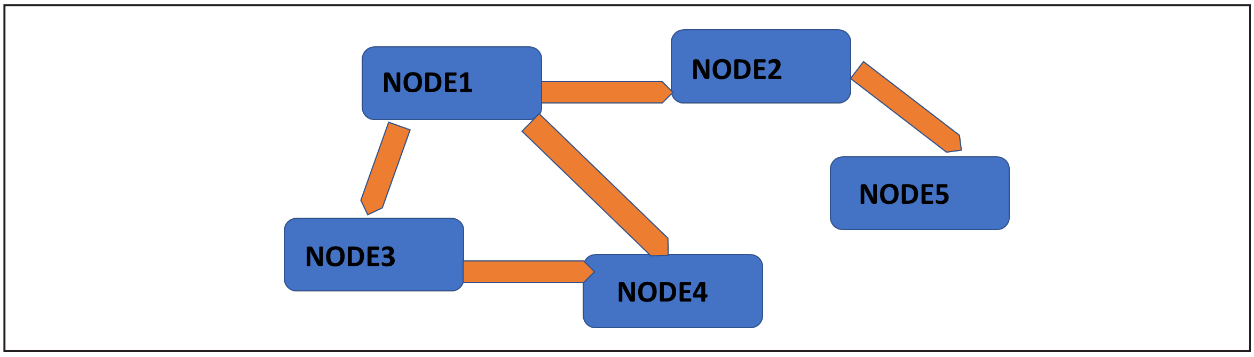

Figure 3. *Sample network with directed edges*

Given the above simple network of five nodes (Figure 3), you may count the in or out going edges (links) for each individual node.

For example, Node1 has three direct connections (to Node2, Node3 and Node4), that is Node1 has 3 degrees. Node5 has only ne direct connection, which goes from Node2, that is Node4 has one degree.

Te degree distribution is simply the tally of the degrees: how many nodes has 1 degree, 2 degrees, etc...

In the above example, the tally of degrees (number of nodes with given number of degrees): 1 degree: 1; 2 degrees: 3; 3 degrees: 1; 4 degrees: 0...

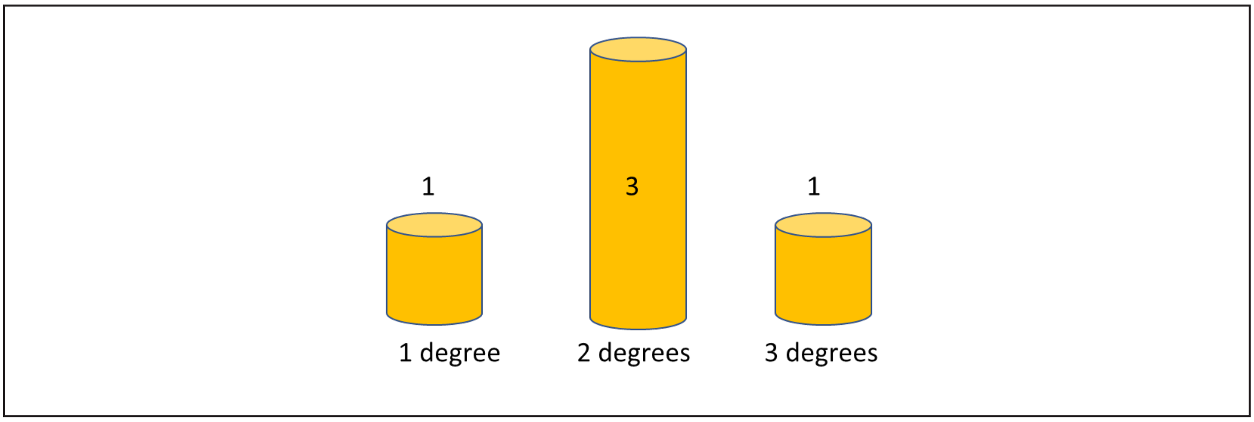

Figure 4. *Degree distribution of the sample network*

When putting the number of nodes on the graph of the degree distribution, you see the number of degrees on the x axis. In this case 3 is our highest degree. The y-axis is the number of nodes with that degree. In our case it is 1 for 3 degrees, 3 for 2 degrees and 1 for 1 degree.

There is a way also to differentiate between in and out degrees in a directed graph. If you count the incoming and outgoing links of each node you will get the degree distribution separately for in and out degrees.

In the above example, the tally of in-degrees (number of nodes with given number of in-degrees): 1 indegree: 3; 2 indegrees: 1; 3 indegrees: 0

Also, the tally of out-degrees (number of nodes with given number of out-degrees): 1 outdegree: 2; 2 outdegrees: 0; 3 outdegrees: 1

It is because 1 indegree you can see in case of Node2, Node3, Node5, 1 outdegree is for Node2 and Node3, etc.

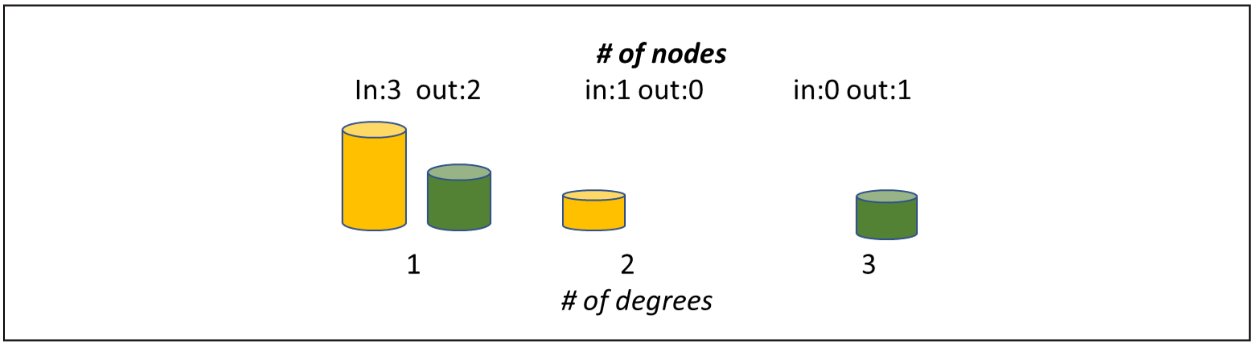

Figure 5. *In and out degree distribution of sample network*

As you can see, there is no nodes with an outdegree of 2 or indegree of 3. But there are 3 nodes with 1 indegree and 2 nodes with 1 outdegree. Also there is 1 node with 3 outdegree.

This is a very simple example. In a network like let's say for example Facebook theer would be a wide range of degrees – as there are people with tens of thousands of friends. In such a complex case the the x-axis for the degree distribution would start also at 0 because there are people with no friends and go up to say a hundred thousand, if there were a case that someone could have that many friends.

Below there is an example of a network of high number of nodes: many modes with very low number of connections (degrees), while very low number of nodes with really huge number of connections.

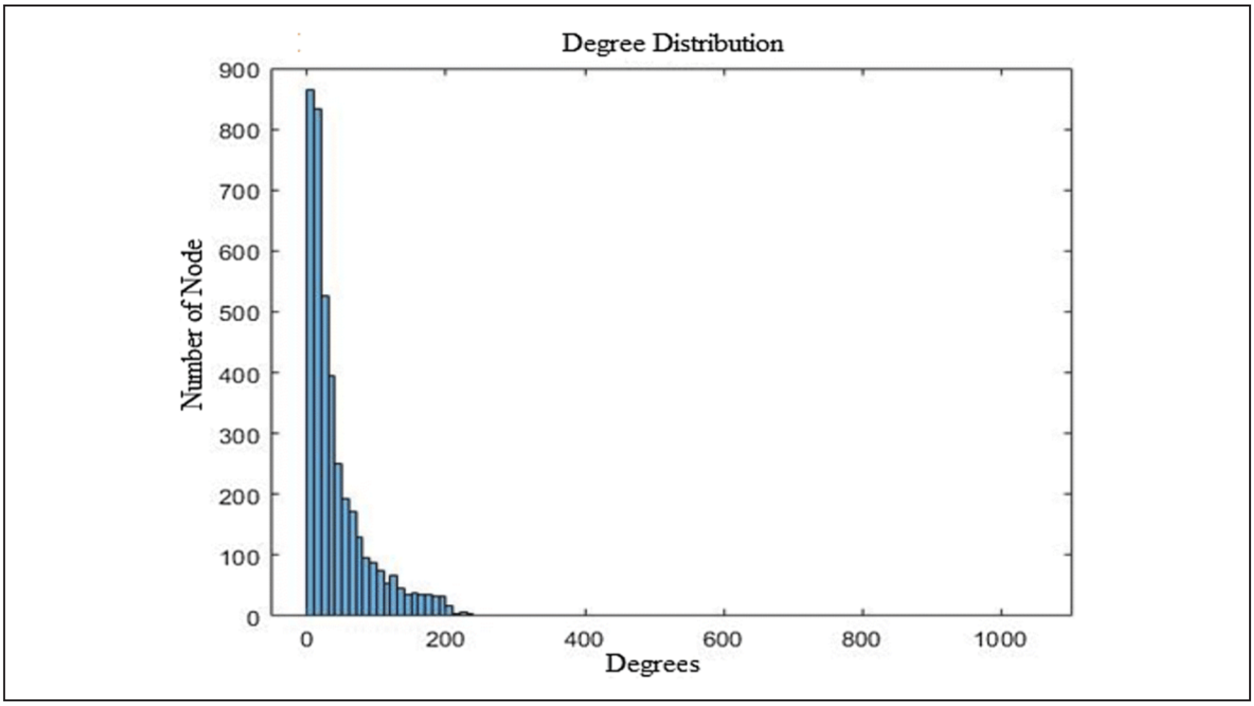

Figure 6. *The power law distribution of networks*

This distribution is called a power-law distribution; unlike the normal distribution it peaks at low x values. Networks tend to follow a power law distribution (instead of a bell curve). Compare the two graphs:

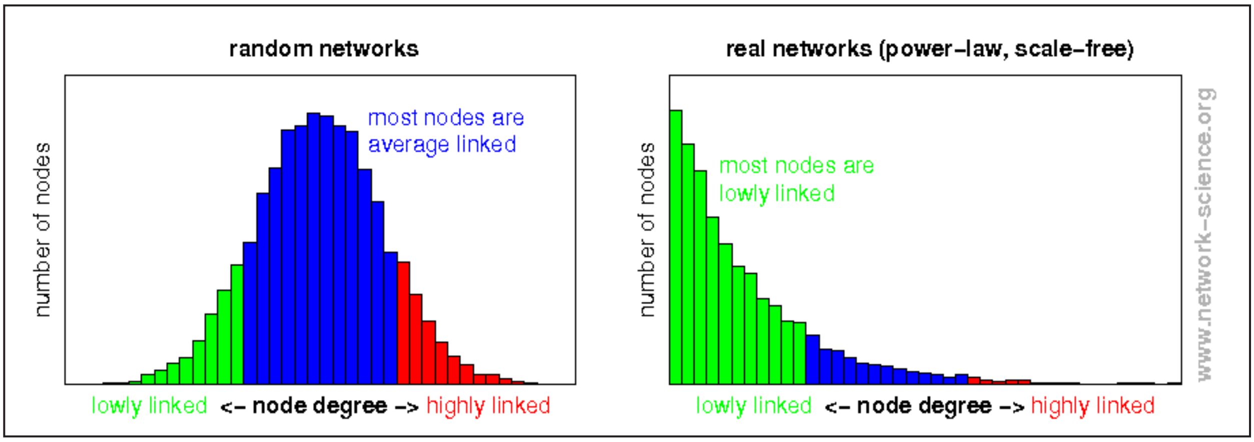

Figure 7. *Random vs. real networks distribution*
Source: www.network-science.org

In networks called random networks with normal degree distribution, most nodes are average linked, but in case of so called real networks, most nodes are low linked.

Besides social networks, book sales is a good example for power law distribution. A lot of books have very low sales but as the sales get higher the number of books with those sales gets lower. So there are definitely books out there that have really high sales but most books are in the lower portion.

7.2.2 Actors and links in agribusiness value chain analysis

Now we understand nodes and edges and degrees of nodes displaying the aount of connections between them. Let's see some of the useful techniques that describes the way how a network is connected and most importantly what these pieces of information on connectedness mean for a practicioner.

Density
Another way of understanding a network is by its density. Density is essentially measuring how many edges are there versus how many edges could there possibly be. In the below example the same network is used with its 5 given edges, but futher 5 possible edges (blue lines) are added to the network.

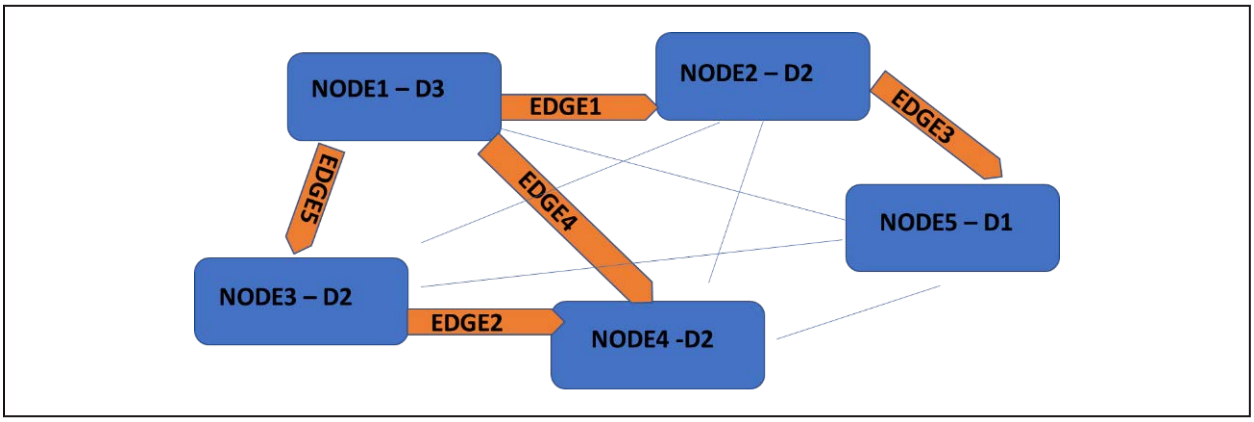

Figure 8. *Density of a network*

So to compute the density we need to know how many edges there are: in this case 5. The number of possible edges – how many possible edges are there if every node were connected to every other node can be defined by the formula of total links.

1 node may have connections to further 4 nodes, which for the case of five nodes gives 4 times 5 connections. This number has to be divided by two, as a link between Node 1 and Node 2 is the same as the link between Node 2 and Node 1.

In the above case we can count five edges that exist and we can count five additional edges: so 10 total edges. The density is then 5 over 10 that is 0.5.

In a realistic network, the densitiy is a low number, normally it is lower than 0.1 or 10%, or even much lower (Facebook has a density of 0.0001). The bigger the network is, the lower the density of it.

Clustering coefficient
There's another way that we can use density to understand networks. The clustering coefficient is a measure to see how well the rest of the network is connected if we remove a certain node. In the following example we take out the one node, let it be Node3 and its connections: Edges5 and 2. What remains is the following.

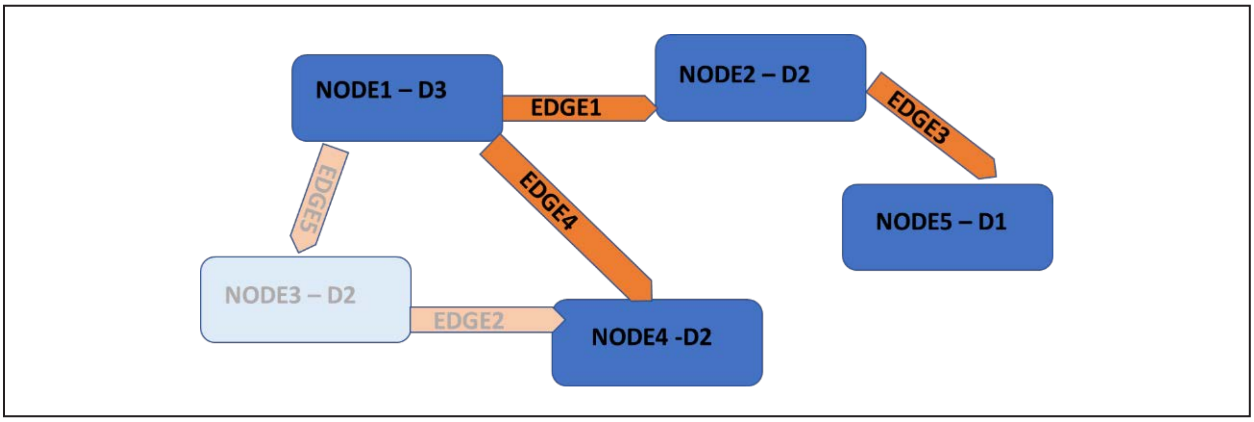

Figure 9. *Remaining part of model network if excluding Node3 and its edges*

We have to calculate only the nodes and connections without Node3. It is easy to admit, that the bigger share of connections remain the less important the removed node was.

Importance of nodes

The way of understanding the importance of nodes in a network gives further information on the network. Let's see the following question: which node is supposed to be considered as most important in this network.

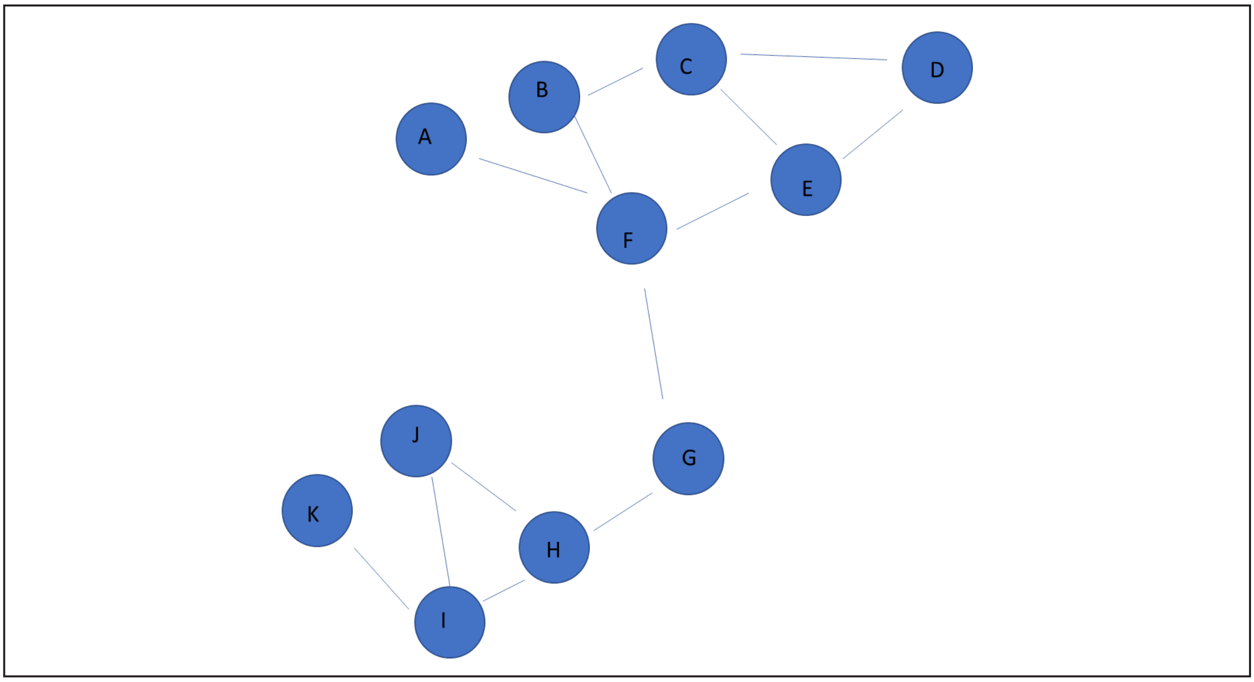

Figure 10. *Sample network of 11 nodes*

Most obvious answers could be Node F or Node G. Node G because it links two big parts of the network and Node F as it has a high number of connections.

Both intuition is right, they are central nodes from different aspects. There are centrality measures that support certain arguments on importance of nodes. When looking at a network, we need to know what each centrality measure means, what is good for measuring and then to be able to make an argument why that centrality measure is the most appropriate for the target of the analysis.

Centrality measures

Centrality is a way of measuring the importance of nodes in a network and there's a variety of ways of doing that. It is possible that they show different nodes to be more important.

Closeness centrality

One of the easiest measures to understand is closeness centrality, which is just the average of the shortest path lengths from one node to every other node in the network. Looking at the above example, let's choose Node F for this exercise. Node A and Node B are 1 shortest path lengths from Node F, as they are directly linked to it. To get to Node C the shothest path takes through either Node B or Node E, therefore the shortest path from Node F to Node C is 2. The following table lists all the shortest path lengths from Node F to every other Node in the network.

Table 1. *Illustration of shortest paths lengths from NodeF*

Node	shortest path length from Node F
A	1
B	1
C	2
D	2
E	1
G	1
H	2
I	3
J	3
K	4

Closeness centrality is simply the average of these shortest path lengths: the sum of the shortest path lengths is 1+1+....3+3=20, to get the average divide it with 10 (being the number of the nodes except for Node F), thus the closeness centrality for Node F is 2.

By calculating the same measure, we find that all the other nodes have higher closeness centrality (CC) values:

CC (F) = 2.0

CC (H) = 2.4

CC (D) = 3.2

....

From the aspect of closeness centrality the most important node in this network is Node F as it is the *closest one to every other node.*

Closeness centrality is really designed to be a centrality measure that looks at how close a node is to every other node in the network. It doesn't measure how big its degree is, it just informs us that it's closely connected to a lot of other nodes. Thios sort of information can be really important for example if we're looking at how diseases or innovation spread.

Degree centrality

Degree centrality is the easiest measure to compute and it is simply the degree of a node. In the above example, according to degree centrality Node F is most central, and Nodes H, I, C and E are coming in second. Nodes A and K are the last with a degree of 1. According to degree centrality what we are really looking for are well connected nodes. So it doesn't matter what role they play in the rest of the network the information on how well connected are they to other people is important.

Betweenness centrality

Betweenness the centrality is one of the most widely used centrality measures when analyzing social networks. The basic intuition is that Betweenness the centrality gives the percentage of shortest paths that include a given note. Let's take the example of Node F and calculate how many of the shortest paths include this particular node.

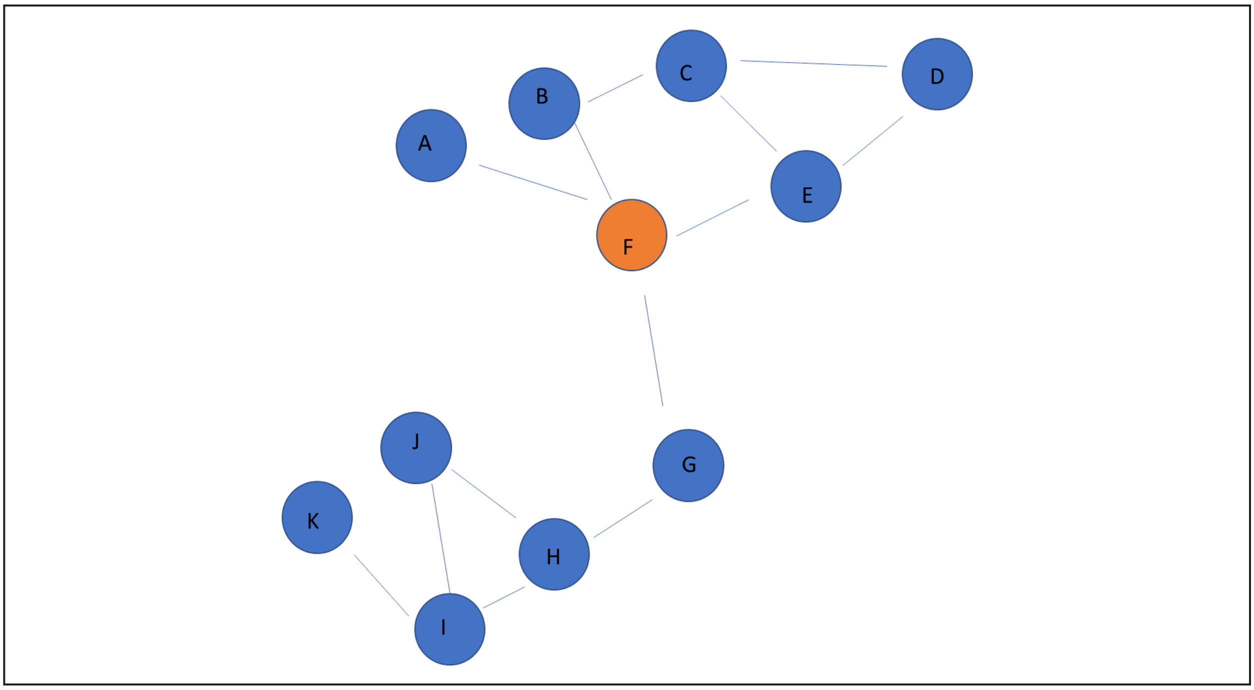

Figure 11. *Illustration of betweenness centrality*

To do so, there is a table indicating all the possible pairs of Nodes and if the shortest path between them includes Node F or not.

Table 2. *Possible pairs of nodes reaching NodeF*

Node from	Node to	includes F?
A	B	0
A	C	0
...		
A	I	1
A	J	1
...		
B	I	1
....		
...		
I	K	0

The total number of shortest paths including Node F is 25 and 10 shorthest paths do not include Node F. So the number of betweenness centrality is 25 over 35, which is 0.71 for Node F.

Doing the same procedure for Node H, the betweenness centrality would be 0.4.

Even in such a simple model it is hard to do computation all the way long, as there is a high number of pairs even in case of eleven nodes.

In a lot of networks there may be a hundred shortest paths. Any network analysis tool will compute betweenness centrality of course.

To summarise it, betweenness centrality measures the degree to which a node is a gatekeeper in the network. So if information is spreading from these nodes over here to these nodes over here particular nodes are critically important if they stop participating in passing on information then no information flows through anymore.

So nodes with high betweenness centrality tend to be really important for connecting different groups and monitoring or helping the flow of information or diseases or other things through networks.

Connectivity measures

Moving on from centrality measures another way of understanding a network is how well connected it is. Connectivity and cohesion measure the minimum number of nodes to remove before the network becomes disconnected.

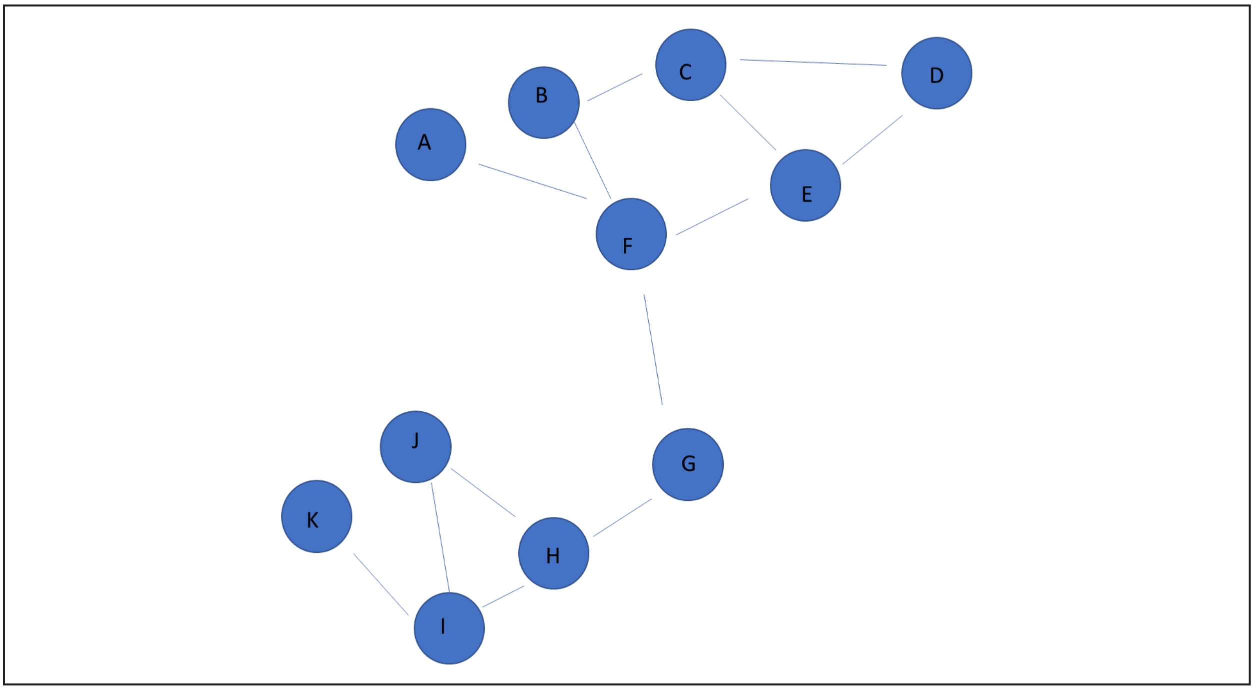

Figure 12. *Sample network for illustration of connectivity measures*

Removing either Node F or Node G the network gets disconnected. Therefore the connectivity measure for this particular network is 1.

Small worlds

In networks of social connections, the average distance to reach out to a specific node from an another one is quite small. Because of the existence of nodes with he number of inks (called as hubs), there are shortest paths going through these nodes that link further nodes. The term for such networks is small world. There is two major properties of small world.

One is that they have a *high average clustering coefficient.* In a small world network nodes' friends tend to know an another more than they would randomly. On the other hand, the *average shortest path length for the network tends to be very short.* That means that people who are in different social circles tend to have people that are connecting them to different groups. So we can get from one point in the network to another quite easily.

This is really interesting structural attribute. Most social networks, but neural networks, the power grids also tend to have this pattern.

Random graphs vs. regular graphs

Unlike in small worlds, random links between nodes form a random graph. The degree distribution follows normal curve. The complement of a random graph is a regular graph.

Random graphs are something that were studied extensively by Paul Erdos who we mentioned in the original presentation in terms of the Erdos number.

He studied a lot of things with graph theory and random graphs and so this is an example of a random graph here. We have a bunch of nodes and you can see there's no real pattern to how the edges appear they're. The complement of a random graph is a regular graph. We can have a regular graph and a random graph with the same number of nodes and same number of edges like these two but they look very different.

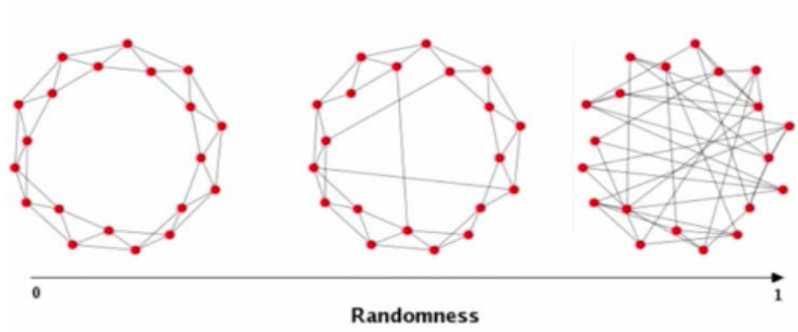

Figure 13. Regular vs. random networks

For a random graph the average shortest path length is very small. You can go from any node to any other pretty quickly because there's a lot of edges that cut across the network making it pretty quick to get from one place to another. So the average shortest path link for the network is short. On the other hand, for a regular graph the average shortest path length is long because if you want to get from a node at the bottom to a node at the top you basically need to go all the way around to get there. If the size triples, the average shortest path length also triples. That's not true for a random graph. On a random graph the size of the average shortest path length increases logarithmically with the size of the graph. On a regular graph it increases linearly. However, on a regular graph there's a generally high clustering coefficient. In a random graph the clustering coefficient tends to be very small.

By combining random and regular graphs' features small world networks can be generated that have both of these features. Removing few edges and rewiring them has very small impact on the clustering coefficient of each node overall. It may decrease it a little bit but not much. On the other hand, those edges that cut across the network makes average shortest path length dramatically smaller[4].

7.3 Networks of agrifood chains

A supply-chain network (SCN) is an evolution of the basic supply chain. Due to rapid technological advancement, organisations with a basic supply chain can develop this chain into a more complex structure involving a higher level of interdependence and connectivity between more organisations, this constitutes a supply-chain network[5].

Often organisations focus only on their organisation; what they produce or provide and not what the end customer receives. Looking at a supply chain network enables firms to look at the overall movement of materials/information from start to end, allowing organisations to see the value in creating partnerships; and the value in working together to ensure the best possible value is provided to the end-customer.

Supply chains and supply networks both describe the flow and movement of materials & information, by linking organisations together to serve the end-customer[6].

Let's look at a supply chain example model in apple juice production.

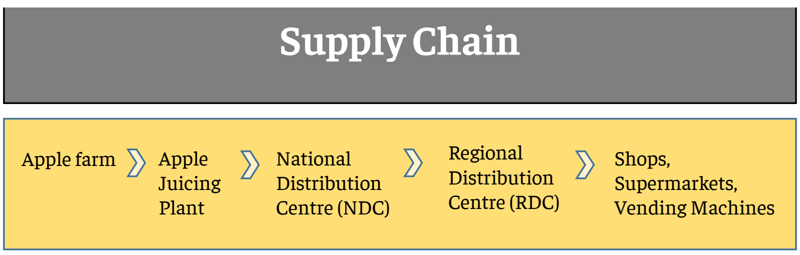

Figure 14. Supply Chain Example: for apple juice production
Source: Hinz[6]

In the above example of an apple juice producer, the flow of materials is seen as a chain of supply from farm to end users.

The apple farm provides fruits for juice production, which entering the distribution chain goes through the regional – local levels of logistics. In the end the juice arrives in shops or to other retailers.

The above diagram is an example of a simplified supply chain. The extended supply chain however includes not only the movement of material flow from the Apple farm through the production process to the end users, but the flow of further materials used in the production. Which pictures the inbound chain of production.

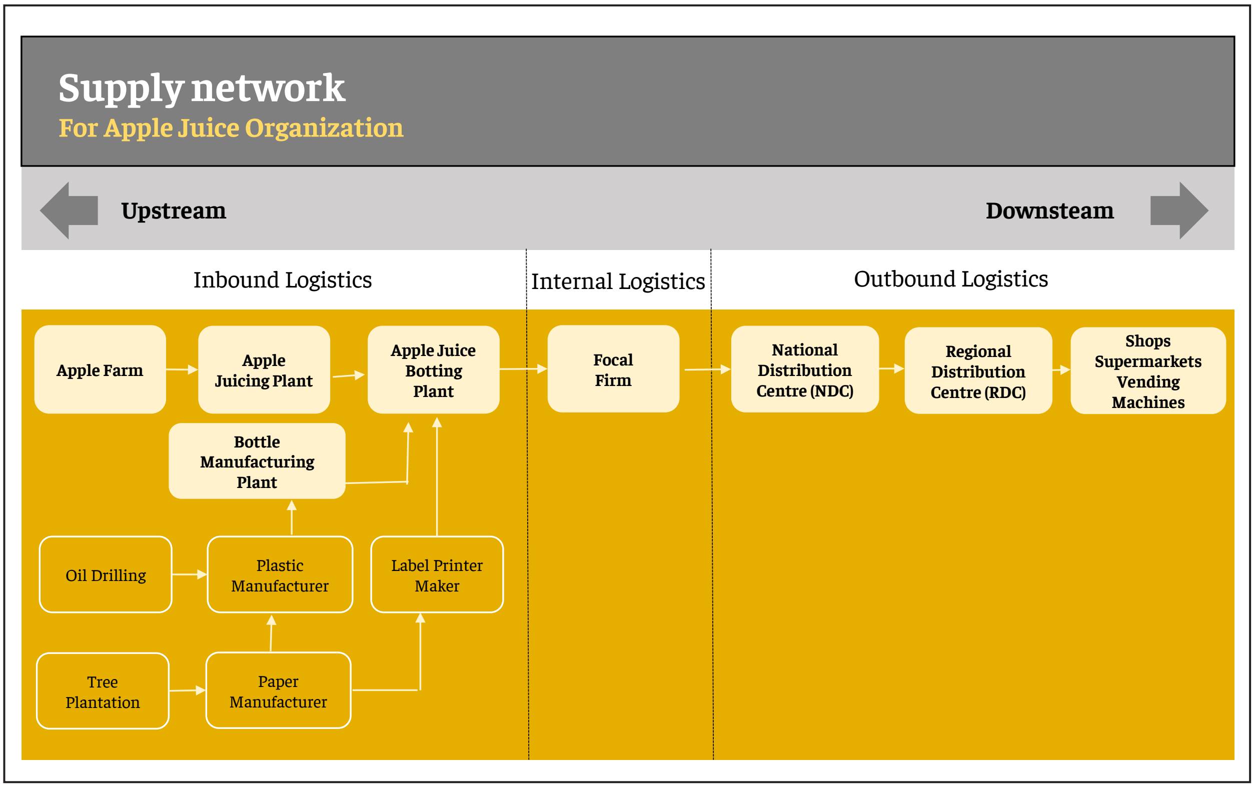

Figure 15. *Extended Supply Chain Example for apple juice production*
Source: Hinz[6]

To get a complete picture of an organisations supply chain network, however, both information and material flows should be mapped. Inefficiency can then be located and removed.

Material flow is the movement of goods from raw primary goods (such as Wool, Trees and Coal etc.) to complete goods (TV's, Radios and Computers) that are to be delivered to the final customer.

Information flow is the demand from the end-customer to preceding organisations in the network. If a focal firm provides their suppliers with their sales data/forecasting demand information; their supplier will be able to reduces costs (such as over production waste) and improve prices. In order to better serve your end customers, it can be important to develop strong partnerships within your supply network which has an effect on flow to end customers, irrespective if being manufacturer, distributor or retailer. Better communication will increase efficiency and productivity. Trust is the core ingredient to develop better communication and relationships.

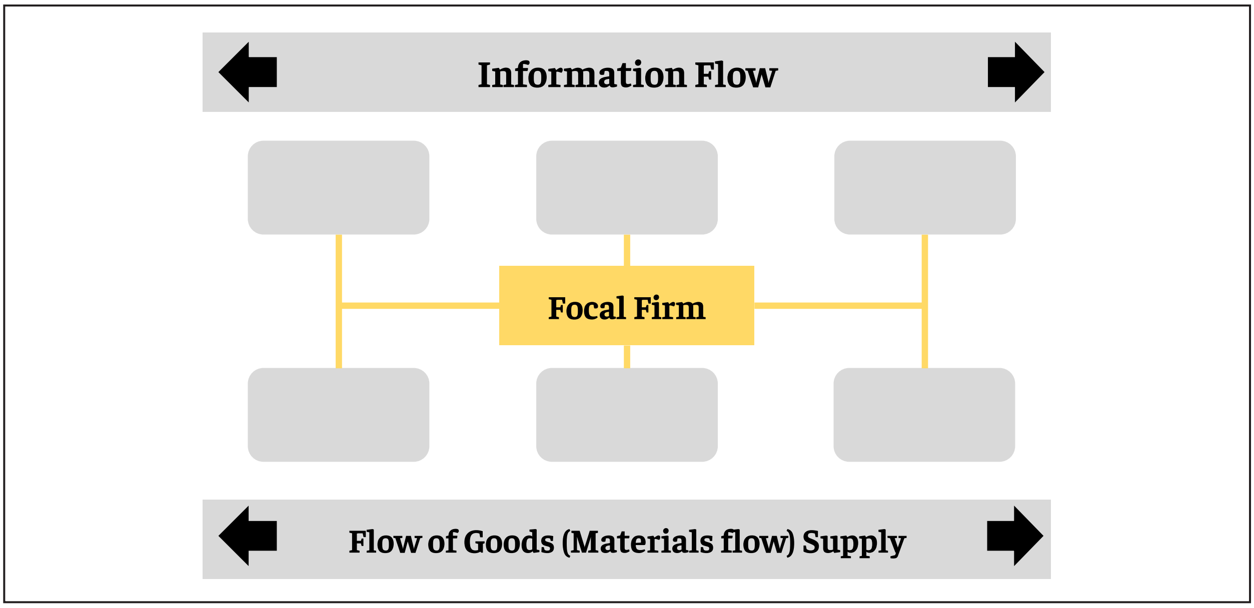

Figure 16. *Information flow vs. material flow simplified model*
Source: Hinz[6]

7.3.1 Case study, presentation of good practice

The following subchapter introduces a model of network of an agricultural producer and service provider company. The idea of seeing the company sales and purchases of a given timeframe in a network provides a new perspective of analysing which products, services, partners, etc. are more vulnerable or valuable for the company. It helps them understanding patterns which enable for example better pricing or better relations management; in general better efficency of the operations.

The overall structure of the company's sales and purchases covers quite a few tables on their products, services, partners, invoices, details of invoices, etc. The data are retrived from the annual accounting records in the period 2007 to 2021.

Table 3. *Elements of data records*

invoice number n = 3228
partner code n = 86
city of partner n = 32
date of financial performance (dd.mm.yyyy) 2007–2021
number of items: 72
number of partners: 6
units of measure
amount of unit
price per unit (HUF)
gross sum of item (HUF)

The network of sales in this company may refer to nodes such as item types, partner codes, or sub items, while the links between the nodes can be defined by the city (in common to the partners), the year of performance (in common to the invoices), or items (in common to the invoices or partners). From these aspects, a possible network can be the nodes of partners (codes), with links representing the city in common, so those partners are linked, who are seated in the same city. The partners (nodes) degree depends on the number of partners seated in the same city. By tallying up the number of partners with various number of degree gives the degree distribution.

In the following example the edges link products and serices (items) that are present in the same invoice, in this way the „streght" of the link (value) represents the number of invoices where these items are jointly listed.

Another representation may be the link (az edges) between invoices (as nodes), when the same item (product or service) is listed in both invoices. In this way the value of the link refers to the number of items the invoices are in common.

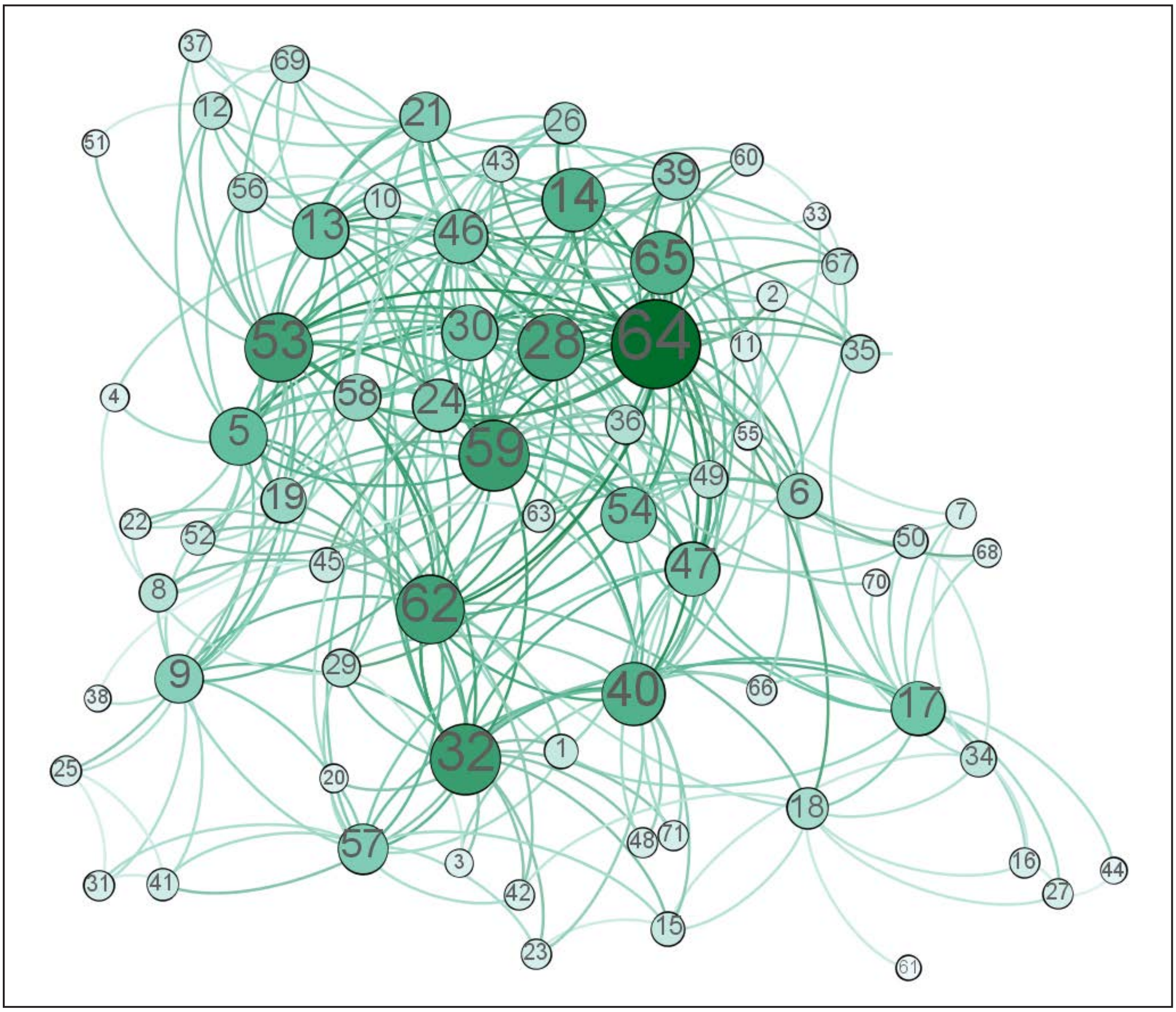

Figure 17. *Network representation of the sample farm's invoices*

In the above network of invoices (Figure 16) there are nodes (products or services) that are listed in many invoices, the highest numbers are 64, 62 and 59. Whilest, some of the items (nodes) are more unique, having only 1, 2 or 4 invoices in common.

Understanding the nature of the network might give an insight on which products or services can be offered as a package by the company, or which of them are individual offers.

Let's introduce the information on the partners. For this purpose, the network illustration is changed and the different colours of edges show those links that belong to different partners.

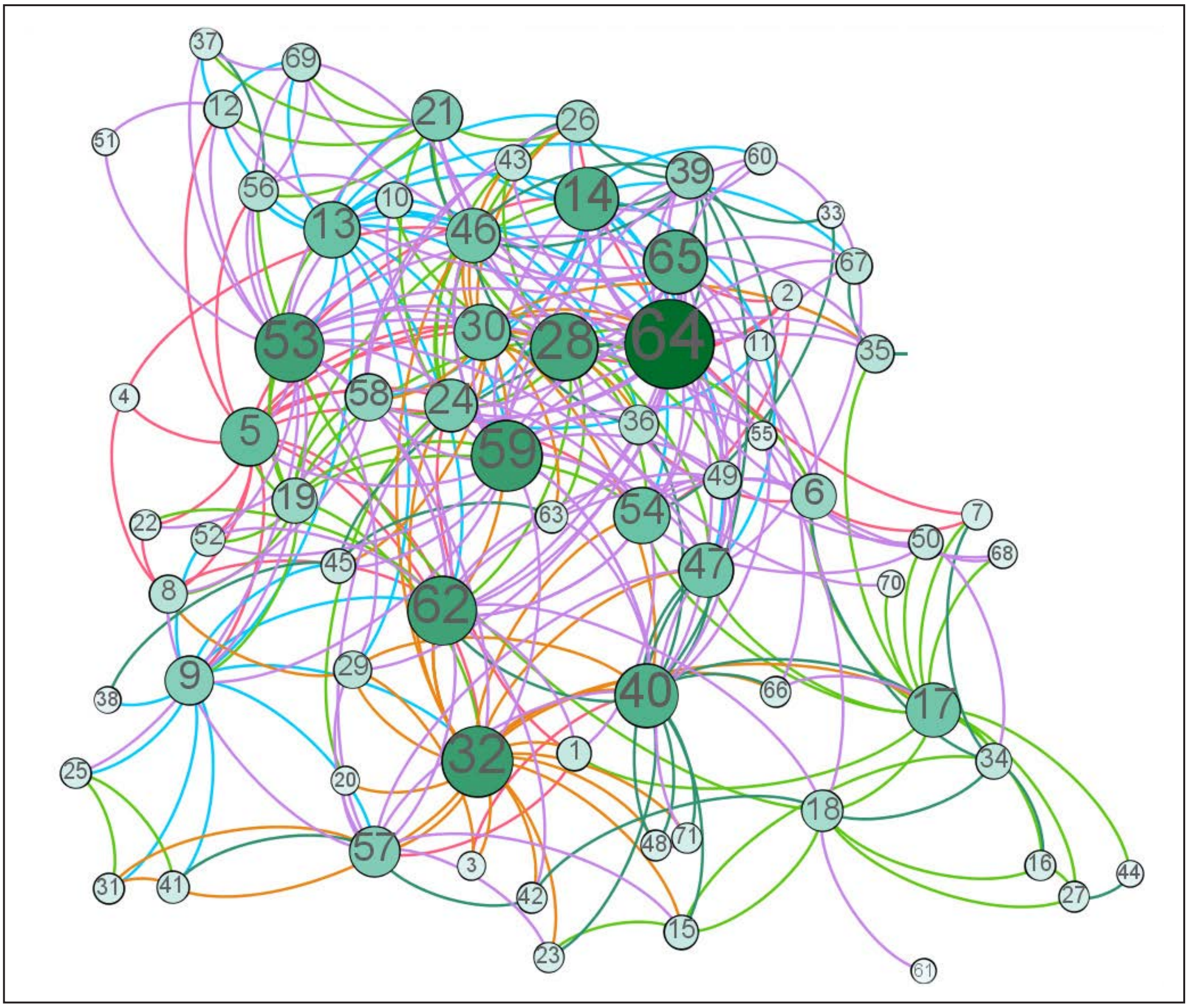

Figure 18. *The farm's partnership structure on invoices and items*

The sample network of the farm' sales invoices illustrates the six partners with differnt colours. By understanding the nature of the network gives information on the differentiation strategy of relations management.

A deeper look at various measures and statitics of the network adds even more details for managers.

Bibliography

[1] Molnár L. (2020) A hálózatelemzés alapfogalmai – gráfok, centralitás, szomszédosság, hidak és a kis világ In: Sasvári, Péter (szerk.) Rendszerelmélet. Ludovika Egyetemi Kiadó Nonprofit Kft. – Ludovika Press, Budapest. pp. 123–140., https://doi.org/10.36250/00734.07
[2] Barabási A. L., Posfai M. (2016) A hálózatok tudománya, Libri Kiadó, Budapest.
[3] Harrrison, A., van Hoek, R. (2011) Logistics Management and Strategy, 3th edition, FT Prentice Hall, Harlow.
[4] Barabási A. L. (2022) Behálózva - A hálózatok új tudománya, Open Books, Budapest.
[5] Slack, N, Chambers S. Harland, C., Harrison, A. and Johnston, R. (1997) Operations Management, Fifth Edition, FT Prentice Hall, Harlow.
[6] Hinz, P. (2011): What is a Supply Chain Network? https://www.adaptalift.com.au/blog/2011-09-27-what-is-a-supply-chain-network Letöltés dátuma: 2021. 11. 12.