A primer on network analysis for business

Clément Levallois

2017-12-08

Table of Contents

last modified: 2018-09-10

[EMLyon logo corp] | *EMLyon_logo_corp.png*

1. Definitions

A network is a dataset made of entities and their relations

Scientists use the term "graph" to discuss networks.

[network 1] | *network-1.png*

Figure 1. This is a network

a. Social networks

As users, we are very familiar with one type of networks - social networks:

[facebook][width=150] [twitter][width=150] [weibo][width=150] [instagram][width=150] [snapchat][width=150] [wechat][width=150] [linkedin][width=150]

[colored network] | *colored-network.png*

Figure 2. source: http://www.minanacheva.com/getting-visual-with-facebook-data/

b. Other networks

It is important to realize that networks cover more than relations between humans. For example, it is possible to imagine a network made out of cooking recipes. 2 ingredients are connected if they appear frequently in the same recipes.

Scanning all recipes and their ingredients from a website of cooking recipes, this gives:

[ingredients network] | *ingredients-network.png*

Figure 3. source: http://arxiv.org/abs/1111.3919

Semantic networks are another broad category of networks. The method is the same: we need to find a way to "relate" words in a text, then we get a network. The general idea is the same as in cooking recipes: 2 terms of a text will be connected in the network if they frequently appeared in same paragraphs.

[editorials] | *editorials.png*

Figure 4. source: http://www.nature.com/nature/journal/v463/n7278/full/463157a.html

c. How big can networks be?

With a surge in computing power in the age of big data, and the adequate NOSQL databases (such as Neo4J or OrientDB), we can deal with huge networks:

For example, "The Anatomy of the Facebook Social Graph" (2011)

→ study of 721 million active Facebook users and the 69 billion (!) friendship links connecting them.

A limit is quickly reached in terms of visualization: it is hard to fit millions of nodes on a screen. In the next visualization, we can see a network of 90,000 Swedish speakers and their relations on Twitter. The view is very cluttered.

(open the source for an interactive version)

[swedish] | *swedish.png*

Figure 5. source: http://twittercensus.se/graph2015/

d. How to discuss networks? Some vocabulary

[Terminology] | *Terminology.png*

Figure 6. Terminology

2. Networks: what use for business?

a. Segmentation

If a network is made of entities and their relations, then a segment is a subgroup of entities in the network, which has some cohesion or something in common.

This subgroup of nodes in the network is often called a "**community**".

Detecting communities in a network, also called "clustering", consists in finding nodes that have many connections in common.

This is a mathematical and algorithmic procedure, but it is very simple to understand visually:

[segmentation with community detection in networks] | *segmentation-with-community-detection-in-*

networks.png

Figure 7. segmentation with community detection in networks

b. Finding key players

[Key players visualized by resizing nodes] | *Key-players-visualized-by-resizing-nodes.png*

Figure 8. Key players visualized by resizing nodes

c. Understanding how information spreads

A data science company created "Where does my tweet go", which traces how a given tweet spreads through retweets. The service is now discontinued (Twitter datan was too expensive to buy) but the mechanism can be explained:

[Where Does my Tweet Go by MFGLabs] | *Where-Does-my-Tweet-Go-by-MFGLabs.png*

Figure 9. Where Does my Tweet Go by MFGLabs

d. Identifying patterns - for fraud detection, control or intelligence.

In the following video, we see participants in the money market (short term loans between banks) in Europe. 2 banks are connected if one lends to the other. The pattern of exchanges shifts through years - banks withdraw from the market.

▶ https://www.youtube.com/watch?v=YvauCrHGWYc *(YouTube video)*

Another example: connecting seemingly unrelated measures of business performance with Oracle BI and Linkurious:

▶ https://www.youtube.com/watch?v=KBIZoUikfwo *(YouTube video)*

3. To go further

[golbeck][width=150,link=https://www.amazon.com/Analyzing-Social-Web-Jennifer-Golbeck/dp/0124055311] [nodexl][width=150,link=https://www.amazon.com/Analyzing-Social-Media-Networks-NodeXL/dp/0123822297] [newman][widtht=150,link=https://www.amazon.com/Networks-Introduction-Mark-Newman/dp/0199206651] [barabasi][width=150,link=https://www.amazon.com/Network-Science-Albert-L-e1szl-f3-Barab-e1si/dp/1107076269]

You can also visit my tutorials on Gephi, the leading software to visualize large graphs:

https://seinecle.github.io/gephi-tutorials/

The end

Find references for this lesson, and other lessons, here.

[round portrait mini 150][align="center", role="right"] This course is made by Clement Levallois.

Discover my other courses in data / tech for business: https://www.clementlevallois.net

Or get in touch via Twitter: @seinecle