UNIVERSIDAD POLITÉCNICA DE YUCATÁN

SOCIAL NETWORK ANALYSIS

HOMEWORK 1

Professor:
Ing. Didier GAMBOA

Network Science: Concepts and Applications

Student:
Walter VIVES
DATA 8A

1 Application Fields of Network Science

A NETWORK is, in its simplest form, a collection of points joined together A in pairs by lines. In the jargon of the field the points are referred to as vertices' or nodes and the lines are referred to as edges. Many objects of interest in the physical, biological, and social sciences can be thought of as networks. [2]

Network Science is concerned with understanding and modeling the behavior of real-world net-worked systems and observational data are the starting point for essentially all the developments of the field.[1]

To describe the detailed behavior of a system consisting of hundreds to billions of interacting components, we need a map of the system's wiring diagram. In a social system this would require an accurate list of your friends, your friends' friends, and so on. In the WWW this map tells us which webpages link to each other. In the cell the map corresponds to a de- tailed list of binding interactions and chemical reactions involving genes, proteins, and metabolites. In the past, we lacked the tools to map these networks. It was equally difficult to keep track of the huge amount of data behind them. The Internet revolution, offering effective and fast data sharing methods and cheap digital storage, fundamentally changed our ability to collect, assemble, share, and analyze data pertaining to real networks.[2]

while many disciplines have made the important contributions to network science, the emergence of a new field was partly made possible by data availability, offering accurate maps of networks encountered in different disciplines. These diverse maps allowed network scientists to identify the universal properties of various network characteristics. This universality offers the foundation of the new discipline of network science.[2]

Network science is defined not only by its subject matter, but also by its methodology.[2]

1.1 Technological Networks

Technological networks, the physical infrastructure networks that have grown up over the last century or so and form the backbone of modern technological societies. Perhaps the most celebrated such network and a relatively recent entry in the field is the Internet, the global network of data connections, electrical, optical, and wireless, that links computers and other information systems together. [1]

Examples of technological networks: power grids, transportation networks, delivery and distribution networks, and telephone networks. The Internet is the worldwide network of physical data connections between computers and related devices. The Internet is a packet switched data network, meaning that messages sent over it are broken up into packets, small chunks of data, that are sent separately over the network and reassembled into a complete message again at the other end. The format of the packets follows a standard known as the Internet Protocol (IP) and includes an IP address in each packet that specifies the packet's destination, so that it can be routed correctly across the network.[1]

Table 1: Technological Networks

Network	Nodes	Links	Directed/Indirected	#Nodes	#Links
Power Network: The case of the Western US and ERCOT grids	buses	transmission lines	Indirected	4941	6594
Transportation	Locations	Distance Roads	Indirected	-	-

The power grid refers to the complex network of energy generators, consumers and connecting transmission lines that are involved in the delivery of electricity. The advent of the 'smart grid' has added further complexity to the system due to the introduction of new renewable generation resources and flexible consumer demand response. [6]

1.2 Information Networks

Networks consisting of items of data linked together in some way. Information networks are all, so far as we know, man-made, with perhaps the best known example being the World Wide Web, though many others exist and are worthy of study. particularly citation networks of various kinds. [1]

In addition, there are some networks which could be considered information networks but which also have social aspects to them. Examples include networks of email communications, networks on social-networking websites such as Facebook or Linkedln, and networks of weblogs and online journals. The classification of networks as social networks, information networks, and so forth is a fuzzy one, and there are plenty of examples that, like these, straddle the boundaries. [1]

Table 2: Information Networks

Network	Nodes	Links	Directed/Indirected	#Nodes	#Links
World Wide Web	Web Pages	Hyperlinks	Directed	-	-
Citation Papers	Paper	Citation between papers	Directed	-	-

1.3 Biological Networks

Within the cells inside our bodies, special molecules called proteins interact in a variety of ways. For example, when a protein folds, its change in structure can regulate the function of another protein or the activity of an enzyme. Enzymes (themselves proteins) catalyze biochemical reactions and are vital to metabolism, which maintains life by harvesting energy for building and supporting the proteins that make up our tissues and organs. Proteins also regulate cell signaling and immune responses. All of these interactions can be seen as networks: protein interaction networks, metabolic networks, gene regulatory networks, and so on. These biological networks exist within a cell. At a higher level, within a body, connections between neural cells (synapses) give rise to the neural networks that form our brains. And at an even higher level, entire species interact. An animal of one species may see

another species as food, creating an ecological network, or food web among species. When we think of this network, ecological balance depends on the availability of species that sustain each other. Removing a node in such a food web — when a species goes extinct, for example — affects the survival of other parts of the ecosystem network. [4]

Table 3: Biological Networks

Network	Nodes	Links	Directed/Indirected	#Nodes	#Links
Protein-protein interaction	Proteins	Protein interactions	Indirected	-	-
Metabolic	Metabolites & Enzymes	Metabolism	Directed	-	-

Protein interaction represents the physical relationships between proteins. They are central to practically every process that takes place in the cell. [5]

Metabolic Networks represent the biochemical relation that allow an organism grow, reproduce, respond to the environment and maintain its structure. Metabolites and enzymes take the role of nodes and the reactions describing their transformations are represented as directed edges. Edges can represent the direction of the metabolic flow or regulatory effects of a specific relation. [5]

1.4 Social Networks

A social Network is a group of people connected bt some type of relationship, friendship collaboration, romance or mere acquaintance are all examples of social relationships that connect pairs of people. When we talk about a social network, er typically think of a particular type of relationship. A person is represented by a node in the social network, and the relationship is represented by a link between two people. The network is therefore a representation of relationship. It allows us to talk about the relationship, to describe it and analyze it at a level that goes beyond a pair of people.[4]

The applications of online social networks have dramatically changed the world. The success of the recent democratic revolution in Egypt was based in part on Twitter and serves as an example to show the impact of a social network running on top of state-of-the-art technological networks. This introduced new insight into collective intelligence or crowd intelligence. Social ants, bees, and schools of fishes, based on very limited local information, can complete complex tasks. Wisdom of the crowd, embodied is such applications as Wikipedia, illustrates the interplay between social networks and technological networks.[3]

Table 4: Social Networks

Network	Nodes	Links	Directed/Indirected	#Nodes	#Links
Data 8A Students	Students	Friendship	Directed	30	-
Facebook	Users	Friendship	Indirected	2.85B approximately	-

2 Write about a network you are personally interested in.

Video games players Network. Let's take as example the game Clash Royale. It is a freemium real-time strategy video game developed and published by Supercell. The game was released globally on March 2, 2016. Clash Royale reached 1 billion in revenue in less than a year on the market. In three years, Clash Royale' made 2.5 billion in revenue according to market intelligence company Sensor Tower. [7]

This company has good earnings, therefore it must be analyzed its data in detail to improve the performance of the game. It would be cool analyze the network of top players. The Matches won between them, in order to determine who is the best one and their location.

1. **What are its nodes and links?** The nodes would be the top players, they will be determine given their maximum number of trophies. You get trophies winning matches. The edges will be the battle between them and the direction will be determine if they win the battle. Therefore, it can be possible that we have a mixed graph.

2. **How large is it?** Clash Royale has approximately 1 million active players. Top players are the best 100 players.

3. **Can be mapped out? Why?** Yes, the network can be mapped. We can see their distribution in the global map, given the location of the players. Then, it can be determined the relations between the players given their location and their victory.

4. **Why do you care about it?** The video game industry has been growing really fast over the years. A network related with this industry can help companies to take better decisions.

   This game is an strategy game that I have been playing since its released on 2016. I like to played it occasionally. Something interesting is that SUPERCELL company has a small group of workers. I was checking their team and they have data scientist and if you want to be part of the team, you have to pass a lot of filters, and you have to have years of experience.

3 Make a brainstorm and answer, what would be the area where network science could have the biggest impact in the next decade? Explain your answer.

- Social Network
- Healthcare
  - vaccine creation
  - metabolic process
  - protein process
  - Diseases like cancer or HPV
  - longevity
- Economics
- blockchain
- autonomous cars
- aerospace
- agroecology

Network Science is present in many areas. Therefore, it is a complex task determine in which one will have the biggest impact. I could say that in the health area will have a big impact given the desire of the human being to have a longer life, technological advances will help a lot to achieve this. Also, in the aerospace area given that the private sector is present in this area.

4 Discuss some methodologies that can address the grand challenges of Social Network Analysis.

1. **Big Data Paradox** Social media data is undoubtedly big. However, when we zoom into individuals for whom, for example, we would like to make relevant recommendations, we often have little data for each specific individual. We have to exploit the characteristics of social media and use its multidimensional, multisource, and multisite data to aggregate information with sufficient statistics for effective mining.[10]

   Big data is all the rage. Its proponents tout the use of sophisticated analytics to mine large data sets for insight as the solution to many of our society's problems. These big data evangelists insist that data-driven decision-making can now give us better predictions in areas ranging from college admissions to dating to hiring. And it might one day help us better conserve precious resources, track and cure lethal diseases, and make our lives vastly safer and more efficient. Big data is not just for corporations. Smartphones and wearable sensors enable believers in the "Quantified Self" to measure their lives in order to improve sleep, lose weight, and get fitter. And recent revelations about the National Security Agency's efforts to collect a database of all caller records suggest that big data may hold the answer to keeping us safe from terrorism as well. [8]

   Big data paradoxes helps us to see the big picture of the big data, what big data is about. For example, the power paradox tell us about how powerful the data can be and how the people in the power can take control of events in real-life, such as presidential elections, avoiding demonstrations by censoring the people in charge, etc. The people in power are usually big business and the government. Basically, someone with power can take advantage of the data.

2. **Obtaining Sufficient Samples** One of the commonly used methods to collect data is via application programming interfaces (APIs) from social media sites. Only a limited amount of data can be obtained daily. Without knowing the population's distribution, how can we know that

our samples are reliable representatives of the full data? Consequently, how can we ensure that our findings obtained from social media mining are any indication of true patterns that can benefit our research or business development? [10]

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. The data collection component of research is common to all fields of study including physical and social sciences, humanities, business, etc. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same. The goal for all data collection is to capture quality evidence that then translates to rich data analysis and allows the building of a convincing and credible answer to questions that have been posed. Regardless of the field of study or preference for defining data (quantitative, qualitative), accurate data collection is essential to maintaining the integrity of research. Both the selection of appropriate data collection instruments (existing, modified, or newly developed) and clearly delineated instructions for their correct use reduce the likelihood of errors occurring.[9]

Data collection is one of the most important stages in conducting a research. You can have the best research design in the world but if you cannot collect the required data you will be not be able to complete your project. Data collection is a very demanding job which needs thorough planning, hard work, patience, perseverance and more to be able to complete the task successfully. Data collection starts with determining what kind of data required followed by the selection of a sample from a certain population. After that, you need to use a certain instrument to collect the data from the selected sample.[9]

Therefore, having enough data as well as having quality data is the basis to be able to develop any project in which we are interested. If the data collected is insufficient or of poor quality, the development of the project will be compromised. The data helps us understand behaviors of n things given its field of study, it also helps us predict future behaviors. Obtaining meaningful data is an important part of the development of any project. The data helps both in the academic field

and in the industrial or government field.

3. **Noise Removal Fallacy** In classic data mining literature, a successful data mining exercise entails extensive data preprocessing and noise removal as "garbage in and garbage out." By its nature, social media data can contain a large portion of noisy data. We have observed two important principles: blindly removing noise can worsen the problem stated in the big data paradox because the removal can also eliminate valuable information, and the definition of noise becomes complicated and relative because it is dependent on our task at hand. [10]

4. **Evaluation Dilemma** A standard procedure of evaluating patterns in data mining is to have some kind of ground truth. For example, a dataset can be divided into training and test sets. Only the training data is used in learning, and the test data serves as ground truth for testing. However, ground truth is often not available in social media mining. Evaluating patterns from social media mining poses a seemingly insurmountable challenge. On the other hand, without credible evaluation, how can we guarantee the validity of the patterns? [10]

5 What marketing opportunities do you think exist in Social Media? Can you outline an example of such an opportunity in given Social Network (Twitter, Facebook, Instagram, Tumblr)?

I consider that Facebook is the social network that has had a good management of the social network, has known how to monetize it. With Facebook Business you can create advertising campaigns all over the world, its platform gives the option to publish ads on Facebook and Instagram. Facebook Business allows you to create campaigns, has an editor, gives you the option of segmenting the market, pausing your advertising campaigns, estimating spending per campaign. I have been learning about the use of said platform and really people invest large amounts of money and the return on investment is greater. If you know how to use the platform and segment your population, it is a win-win. This is precisely why I consider that Facebook together with Instagram are the best social network for digital marketing.

For example, if you have an art business that sells paintings, the ideal is to create an ad campaign, create the ads, and then divide the ad groups. In each ad you can edit the content, you can decide in which part of Facebook you want it to appear, for example in the Feed, in the side columns (if you want it to be displayed in the desktop version), the ad can even appear on Facebook Messenger. Also an important part in the development of advertising campaigns is to have a work plan defined, Facebook gives you the option to select what is your goal to achieve with advertising campaigns, it can be to gain new followers for the page, receive comments, messages, reactions and so on. Segmenting the population is important because if you don't do it, you are probably burning your money because the ads are going to be shown to people who are not interested in your product. Knowing how to handle Facebook Business Manager has its certain level of difficulty, it is not impossible. The good side of this is that the investment is worth it, it is a win-win.

6 Consider the road map in Figure 1. If one were creating a network representation of traffic patterns, which of the following would be the best choice to make up the links of the network? (Hint: your answer to the next question may inform your answer to this question, and vice-versa.)

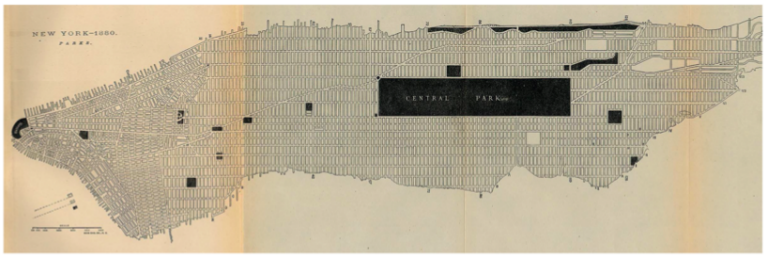

Figure 1: Map of New York in 1880. From Report on the Social Statistics of Cities, Compiled by George E. Waring, Jr., United States Census Office, 1886. Image courtesy of University of Texas Libraries.)

1. Pedestrians traveling along the streets.
2. Road segments, e.g., 5th Ave. between 12th and 13th streets.
3. Entire roads, e.g., 5th Ave.
4. **Vehicles traveling on the roads.**

I consider that the links have to be the **vehicles traveling on the roads** given that if we want to know traffic patterns we must know the frequency in which vehicles pass through a certain area, in this case the streets of New York City.

7 Consider the road map in Figure 1. In the network representation of traffic patterns, which of the following would be the best choice to make up the nodes of the network? (Hint: your answer to the previous question may inform your answer to this question, and vice-versa.)

1. City blocks, e.g., the block between 5th-6th avenues and 12th-13th streets.
2. **Street intersections, e.g., 5th Ave. and 12th St.**
3. Pedestrians moving along the streets
4. Vehicles traveling on the roads

In the previous question I considered that the links of the traffic network must be the Vehicles traveling on the roads because we want to measure the frequency to determine that in which area there is more vehicular flow, therefore our nodes must be the **street intersections**, so we can determine the vehicular flow between the streets and with this determine in which street there is more traffic according to the vehicular frequency in the area.

8 Compare the US air transportation network in Figure 2 with the Manhattan road map in Figure 1. The air transportation network displays a distinguishing feature that the Manhattan road network lacks. What is this key characteristic?

1. Singleton nodes with no links.
2. Multiple routes between nodes.
3. Nodes with more than one connected link.
4. **Hub nodes with many links.**

The characteristic that figure 2 has and figure 1 does not, is that figure 2 has **hub nodes**.

This is because figure 2 represents the US transportation network, the nodes are the airports and the edges are the connections between the different airports. In the case of figure 1 we have the map of New York, the nodes are the intersections between streets and the edges are Vehicles traveling on the roads. In theory each node must have 4 edges because the node is the intersection of the streets, therefore this network would lack Hub nodes. Hub is a node with a large degree, meaning it has connections with many other nodes. In the case of figure 2 there may be nodes with many connections given that each airport may have many requests for loading and/or unloading of products.

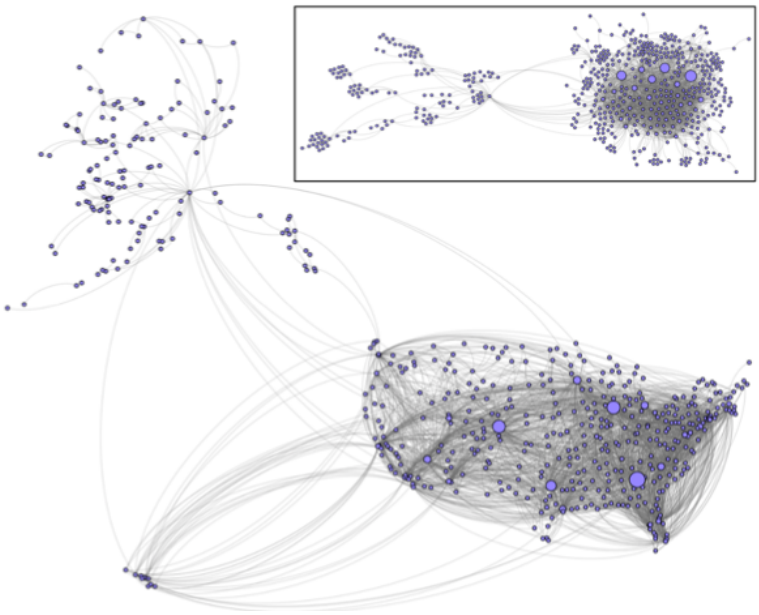

Figure 2: The US air transportation network(flight data from OpenFlights.org). Nodes are positioned according to the geographic coordinates of the corresponding airports, so that we can make out the shape of the continental United States, Alaska, and Hawaii. Note that the map projection makes Alaska appear bigger than its actual size due to its latitude. The airpot hubs with most connections (e.g., Atlanta, Chicago, Denver) are clearly recognizable. The inset maps the same network, but with a different "force-directed" layout, discussed in Section 1.10.

References

[1] Newman, M. (2018). Networks. Oxford university press.

[2] Barabási, A. L. (2013). Network science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1987), 20120375.

[3] Chen, K. C., Chiang, M., Poor, H. V. (2013). From technological networks to social networks. IEEE Journal on Selected Areas in Communications, 31(9), 548-572.

[4] Menczer, F., Fortunato, S., Davis, C. A. (2020). A first course in network science. Cambridge University Press.

[5] Network analysis of protein interaction data. An introduction. Available at: https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/types-of-biological-networks/

[6] Deka, D., Vishwanath, S., Baldick, R. (2016). Analytical models for power networks: The case of the western US and ERCOT grids. IEEE Transactions on Smart Grid, 8(6), 2794-2802.

[7] Wikipedia. (2021). Clash Royale. Available at: https://en.wikipedia.org/wiki/Clash_Royale

[8] Ricards, N., King, J. (2013). Three Paradoxes of Big Data. Stanford Law Review. Available at: https://www.stanfordlawreview.org/online/privacy-and-big-data-three-paradoxes-of-big-data/

[9] Kabir, Syed Muhammad. (2016). METHODS OF DATA COLLECTION.

[10] Zafarani, R., Abbasi, M. A., Liu, H. (2014). Social media mining: an introduction. Cambridge University Press.