Data and Networks
Part II

Dr. Franck Kalala M

African Institute for Mathematical Sciences (AIMS) South Africa

franckm@aims.ac.za

May 16, 2017

Course Content

1. Lecture 1: Introduction
2. Lecture 2: Networks and Graphs
3. Lecture 3: Random network
4. Lecture 4: Scale free and Small world network
5. Lecutre 5: Centrality

Softwares

1. Networkx (Pyhton) (https://networkx.github.io/)
2. igraph (Python, C, R) (http://igraph.org/redirect.html)
3. pajet, (for vizualisation)
   http://vlado.fmf.uni-lj.si/pub/networks/pajek/
4. Gephi, (for vizualisation) https://gephi.org/

Network datasets

1. Stanford Large Network Dataset Collection,
   https://snap.stanford.edu/data/.
2. Network Repository, http://networkrepository.com/.
3. Pajek datasets, http://vlado.fmf.uni-lj.si/pub/networks/data/

Data and Networks (Part II)

Choose one topic of your choice

1. Degree correlations
2. Network Robusteness
3. Spreading phenomena in networks
4. Communicability function in networks

Some readings (BOOKS)

1. E. Estrada, The structure of complex networks: theory and applications, 2011, OUP.
2. E. Estrada and Knigth, A first course in network theory, 2015, OUP.
3. M. Newman, Networks: an introduction, 2010, OUP.

ESSAYS

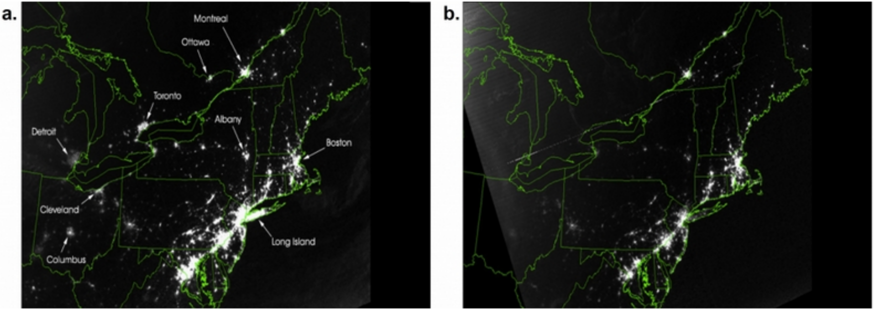
1. Fidele Tubananza (from Rwanda), AIMS-Senegal, 2014
   Network approach of epidemic spreading (stochastic model)
2. Reham Bashir (from Soudan), AIMS-Tanzania,2015
   Centrality in complex networks and application
3. Alice Nayanzi (from Ouganda), AIMS-South Africa, 2016
   Laplacian matrix of a networks and applications, (Research Center)
4. Laeticia Shoma (from DRC), AIMS-South Africa, 2016
   Network spectra and applications.
5. Emily S. Muller (ffrom Sout Africa), AIMS-South Africa, 2017 (current)
   Dynamic Model for social network: The stochastic actor-oriented approach.

$\sum\bigcirc$

You can also google every thing you want.

Vulnerability Due to Interconnectivity

2003 North American Blackout



a. Satellite image on Northeast United States on August 13th, 2003,at 9:29pm (EDT), 20 hours before the 2003 blackout.
b. The same as above, but 5 hours after the blackout.

Vulnerability Due to Interconnectivity

The 2003 blackout is a typical example of a cascading failure.

1. When a network acts as a transportation system, a local failure shifts loads to other nodes.
2. The failure will redistribute to neighboring nodes depending on the size of the load.
3. In no time, we are faced with a cascading event, whose magnitude depends on the position and the capacity of the nodes that failed initially.

Vulnerability Due to Interconnectivity

Cascading failures have been observed in many complex systems.

1. Internet: when traffic is rerouted to bypass malfunctioning routers.
2. Financial systems: when the International Monetary Fund pressured the central banks of several Pacific nations to limit their credit, which defaulted multiple corporations, eventually resulting in stock market crashes worldwide
3. The 2009-2011: financial meltdown is often seen as a classic example of a cascading failure, the US credit crisis paralyzing the economy of the globe, leaving behind scores of failed banks, corporations, and even bankrupt states

Cascading failures can be also induced artificially.

1. An example is the worldwide effort to dry up the money supply of terrorist organizations, aimed at crippling their ability to function.
2. Similarly, cancer researchers aim to induce cascading failures in our cells to kill cancer cells.

Vulnerability Due to Interconnectivity

The Northeast blackout illustrates several important themes in network science:

1. First, to avoid damaging cascades, we must understand the structure of the network on which the cascade propagates.
2. Second, we must be able to model the dynamical processes taking place on these networks, like the flow of electricity.
3. Finally, we need to uncover how the interplay between the network structure and dynamics affects the robustness of the whole system.

☒

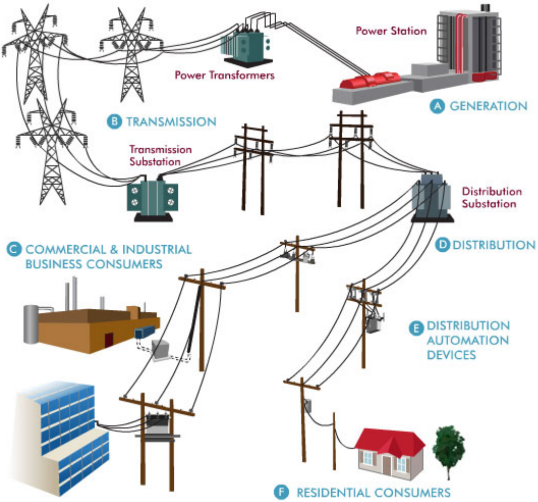
Although cascading failures may appear random and unpredictable, they follow reproducible laws that can be quantified and even predicted using the tools of network science.

Vulnerability Due to Interconnectivity

The blackout also illustrates a bigger theme: vulnerability due to interconnectivity.

1. Indeed, in the early years of electric power each city had its own generators and electric network.
2. therefore, to link neighboring cities up, allowing them to share the extra production and borrow electricity if needed.
3. We owe the low price of electricity today to the power grid, the network that emerged through these pairwise connections, linking all producers and consumers into a single network.
4. It allows cheaply produced power to be instantly transported anywhere.
5. Electricity hence offers a wonderful example of the huge positive impact networks have on our life.

Vulnerability Due to Interconnectivity



local failures, like the breaking of a fuse somewhere in Ohio, may not stay local any longer.

1. Their impact can travel along the network's links and affect other nodes, consumers and individuals apparently removed from the original problem.

In general interconnectivity induces a remarkable non-locality:

spreading

It allows information, memes, business practices, power, energy, obesity, happiness and viruses to spread on their respective social or technological networks, reaching us, no matter our distance from the source.

Hence networks carry both benefits and vulnerabilities.

Uncovering the factors that can enhance the spread of traits deemed positive, and limit others that make networks weak or vulnerable, is one of the goals of network science.

Networks at the Heart of Complex Systems

Stephen Hawking

"I think the next century will be the century of complexity".

We are surrounded by systems that are hopelessly complicated.

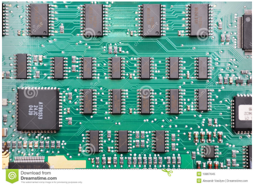
Example

1. society that requires cooperation between billions of individuals, or communications infrastructures that integrate billions of cell phones with computers and satellites.
2. Our ability to reason and comprehend our world requires the coherent activity of billions of neurons in our brain.
3. Our biological existence is rooted in seamless interactions between thousands of genes and metabolites within our cells.

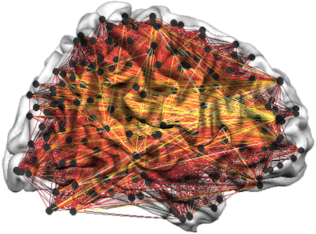
Networks at the Heart of Complex Systems

(a) internet

(b) swarm



(c) electronic system



(d) brain

Networks at the Heart of Complex Systems

complex systems

These systems are collectively called complex systems, capturing the fact that it is difficult to derive their collective behavior from a knowledge of the systems components.

complex systems

Given the important role complex systems play in our daily life, in science and in economy, their understanding, mathematical description, prediction, and eventually control is one of the major intellectual and scientific challenges of the 21st century.

emergence of network science

Υ

The emergence of network science at the dawn of the 21st century is a vivid demonstration that science can live up to this challenge. Indeed, behind each complex system there is an intricate network that encodes the interactions between the system's components:

1. The network encoding the interactions between genes, proteins, and metabolites integrates these components into live cells. The very existence of this cellular network is a prerequisite of life.
2. The wiring diagram capturing the connections between neurons, called the neural network, holds the key to our understanding of how the brain functions and to our consciousness.
3. The sum of all professional, friendship, and family ties, often called the social network, is the fabric of the society and determines the spread of knowledge, behavior and resources.
4. Communication networks, describing which communication devices interact with each other, through wired internet connections or wireless links, are at the heart of the modern communication system.
5. The power grid, a network of generators and transmission lines, supplies with energy virtually all modern technology.
6. Trade networks maintain our ability to exchange goods and services, being responsible for the material prosperity that the world has enjoyed since WWII (Image 1.2).

1. Networks are also at the heart of some of the most revolutionary technologies of the 21st century, empowering everything from Google to Facebook, CISCO, and Twitter.
2. At the end, networks permeate science, technology, business and nature to a much higher degree than it may be evident upon a casual inspection.

※

Consequently, we will never understand complex systems unless we develop a deep understanding of the networks behind them.

Exploding interest in network science

The exploding interest in network science during the first decade of the 21st century is rooted in the discovery that despite the obvious diversity of complex systems,

⋇

the structure and the evolution of the networks behind each system is driven by a common set of fundamental laws and principles.

⋇

Therefore, notwithstanding the amazing differences in form, size, nature, age, and scope of real networks, most networks are driven by common organizing principles.

Exploding interest in network science

⋇

Once we disregard the nature of the components and the precise nature of the interactions between them, the obtained networks are more similar than different from each other.

⋇

What are the forces that have led to the emergence of network science? and what is its impact on science, technology, and society?

Two Forces Helped the Emergence of Network Science

Why didn't we have network science two hundred years earlier?

All many of the networks that the field explores are by no means new:

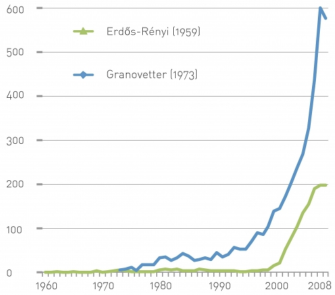
1. metabolic networks date back to the origins of life, with a history of four billion years,
2. social network is as old as humanity.
3. Furthermore, many disciplines, from biochemistry to sociology and brain science, have been dealing with their own networks for decades.
4. Graph theory, a prolific subfield of mathematics, has explored graphs since 1735.

Is there a reason, therefore, to call network science the science of the 21st century?

The Emergence of Network Science

Networks has a long history, with roots in graph theory and sociology, but emerged as a science only during the first decade of the 21st century.

Figure : The explosive interest in networks is well documented by the citation pattern of two classic papers, the 1959 paper by Paul Erdos and Alfred Renyi that marks the beginning of the study of random networks in graph theory and the 1973 paper by Mark Granovetter, the most cited social network paper.



The Emergence of Network Science

Two forces have contributed to the emergence of network science:

1.The Emergence of Network Maps

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.
2. In a social system this would require an accurate list of your friends, your friends friends, and so on.
3. In the WWW this map tells us which webpages link to each other.
4. In the cell the map corresponds to a detailed list of binding interactions and chemical reactions involving genes, proteins, and metabolites.

1.The Emergence of Network Maps

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.

internet evolution

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.

technological advances, at the turn of the millenium allow an explosion of map making.

1. the CAIDA projects that offered the first large-scale maps of the Internet; (CAIDA:Center for Applied Internet Data Analysis),
2. to the hundreds of millions of dollars spent by biologists to experimentally map out protein-protein interactions in human cells;
3. the efforts made by social network companies, like Facebook, Twitter, or LinkedIn, to develop accurate depositories of our friendships and professional ties;
4. the Connectome project of the US National Institute of Health that aims to systematically trace the neural connections in mammalian brains.
5. The sudden availability of these maps at the end of the 20th century has catalyzed the emergence of network science.

2.The Universality of Network Characteristics

It is easy to list the differences between the various networks we encounter in nature or society:

1. the nodes of the metabolic network are tiny molecules and the links are chemical reactions governed by the laws of chemistry and quantum mechanics;
2. the nodes of the WWW are web documents and the links are URLs guaranteed by computer algorithms;
3. the nodes of the social network are individuals and the links represent family, professional, friendship, and acquaintance ties.

2.The Universality of Network Characteristics

the processes that generated these networks also differ greatly:

1. metabolic networks were shaped by billions of years of evolution;
2. the WWW is built by the collective actions of millions of individuals and organizations;
3. social networks are shaped by social norms whose roots go back thousands of years.

Given this diversity in size, nature, scope, history, and evolution, one would not be surprised if the networks behind these systems would differ greatly

The Emergence of Network Science

The Universality of Network Characteristics

A key discovery of network science is that the architecture of networks emerging in various domains of science, nature, and technology are similar to each other, a consequence of being governed by the same organizing principles. Consequently we can use a common set of mathematical tools to explore these systems.

This universality is one of the guiding principle in network science:

1. we will not only seek to uncover specific network properties, but each time we ask how widely they apply.
2. We will also aim to understand their origins, uncovering the laws that shape network evolution and their consequences on network behavior.

The Emergence of Network Science

summary

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.

The Characteristics of Network Science

Network science is defined not only by its subject matter, but also by its methodology. In this section we discuss the key characteristics of the approach network science adopted to understand complex systems.

1. Interdisciplinary Nature
2. Empirical, Data Driven Nature
3. Quantitative and Mathematical Nature
4. Computational Nature

Interdisciplinary Nature

Network science offers a language through which different disciplines can seamlessly interact with each other.

cell biologists, brain scientists and computer scientists alike are faced with the task of characterizing the wiring diagram behind their system, extracting information from incomplete and noisy datasets, and understanding their systems' robustness to failures or attacks.

the common nature of many issues these fields struggle with has led to a cross-disciplinary fertilization of tools and ideas.

1. the concept of betweenness centrality that emerged in the social network literature in the 1970s, today plays a key role in identifying high traffic nodes on the Internet.
2. Similarly algorithms developed by computer scientists for graph partitioning have found novel applications in identifying disease modules in medicine or detecting communities within large social networks.

Interdisciplinary Nature

Mapping the brain

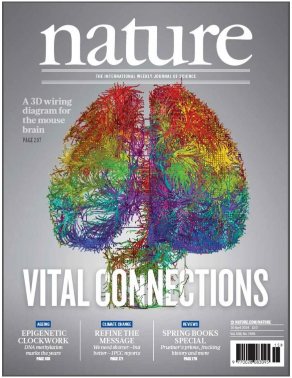


Figure : the April 10, 2014 issue of Nature, reporting an extensive map of the laboratory mouse generated by researchers at the Allen Institute in Seattle.

An exploding application area for network science is brain research. The wiring diagram of a complete nervous system has long been available for C. elegans, a small roundworm, but neuronal connectivity data for larger animals has been missing until recently.

Empirical, Data Driven Nature

Several key concepts of network science have their roots in graph theory, a fertile field of mathematics. Graphy theory will be introduced in the next chapter.

Network science vs Graph theory

What distinguishes network science from graph theory is its empirical nature, i.e. its focus on data, function and utility.

1. In network science we are never satisfied with developing abstract mathematical tools to describe a certain network property.
2. Each tool we develop is tested on real data and its value is judged by the insights it offers about a systems properties and behavior.

Quantitative and Mathematical Nature

To contribute to the development of network science and to properly use its tools, it is essential to master the mathematical formalism behind it.

To contribute to the development of network science and to properly use its tools, it is essential to master the mathematical formalism behind it.

1. Network science borrowed the formalism to deal with graphs from graph theory and the conceptual framework to deal with randomness and seek universal organizing principles from statistical physics.
2. Lately, the field is benefiting from concepts borrowed from engineering, like control and information theory, allowing to understand the control principles of networks, and from statistics, helping us extract information from incomplete and noisy datasets.

Quantitative and Mathematical Nature

$\bowtie$

1. The development of network analysis software has made the tools of network science available to a wider community, even those who may not be familiar with the intellectual foundations and the full mathematical depths of the discipline.
2. Yet, to further the field and to efficiently use its tools, we need to master its theoretical formalism.

Computational Nature

computational challenges

Given the size of many of the networks of practical interest, and the exceptional amount of auxiliary data behind them, network scientists are regularly confronted by a series of formidable computational challenges.

Hence, the field has a strong computational character, actively borrowing from algorithms, database management and data mining. A series of software tools are available to address these computational problems, enabling practitioners with diverse computational skills to analyze the networks of interest to them.

In summary, a mastery of network science requires familiarity with each of these aspects of the field. It is their combination that offers the multi-faceted tools and perspectives necessary to understand the properties of real networks.

Social impacts

The impact of a new research field is measured both by its intellectual achievements as well as by its societal impact, indicated by the reach and the potential of its applications. While network science is a young field, its impact is everywhere.

Economic Impact: From Web Search to Social Networking

The most successful companies of the 21st century, from Google to Facebook, Twitter, LinkedIn, Cisco, Apple and Akamai, base their technology and business model on networks.

The search technology of google is deeply interlinked with the network characteristics of the Web.

Social impacts

Economic Impact: From Web Search to Social Networking

1. Networks have gained particular popularity with the emergence of Facebook, the company with the ambition to map out the social network of the whole planet.
2. Facebook was not the first social networking site and it is likely not the last either: An impressive ecosystem of social networking tools, from Twitter to LinkedIn are fighting for the attention of millions of users.
3. Algorithms conceived by network scientists fuel these sites, aiding everything from friend recommendation to advertising.

Social impacts

Health: From Drug Design to Metabolic Engineering

Completed in 2001, the human genome project offered the first comprehensive list of all human genes [5, 6]. Yet, to fully understand how our cells function, and the origin of disease, a full list of genes is not sufficient:

We also need an accurate map of how genes, proteins, metabolites and other cellular components interact with each other. Indeed, most cellular processes, from food processing to sensing changes in the environment, rely on molecular networks. The breakdown of these networks is responsible for human diseases.

Social impacts

Health: From Drug Design to Metabolic Engineering

1. Network biology
   The increasing awareness of the importance of molecular networks has led to the emergence of network biology, a new subfield of biology that aims to understand the behavior of cellular networks.
2. Network Medecine
   network medicine, aims to uncover the role of networks in human disease.

The importance of these advances is illustrated by the fact that Harvard University in 2012 started the Division of Network Medicine, that employs researchers and medical doctors who apply network-based ideas towards understanding human disease.

Social impacts

Health: From Drug Design to Metabolic Engineering

1. network pharmacology
   Networks play a particularly important role in drug development. The ultimate goal of network pharmacology [7] is to develop drugs that can cure diseases without significant side effects.

This goal is pursued at many levels, from millions of dollars invested to map out cellular networks, to the development of tools and databases to store, curate, and analyze patient and genetic data.

Network Biology and Medicine

Figure : The cover of two issues of Nature Reviews Genetics, the leading review journal in genetics. The journal has devoted exceptional attention to the impact of networks: the 2004 cover focuses on network biology (top), the 2011 cover discusses network medicine [9] (bottom).

Social impacts

Epidemics: from Forecasting to Halting Deadly Viruses

While the H1N1 pandemic was not as devastating as it was feared at the beginning of the outbreak in 2009, it gained a special role in the history of epidemics:

It was the first pandemic whose course and time evolution was accurately predicted months before the pandemic reached its peak. This was possible thanks to fundamental advances in understanding the role of transportation networks in the spread of viruses.

Social impacts

Epidemics: from Forecasting to Halting Deadly Viruses

Before 2000 epidemic modeling was dominated by compartment-based models, assuming that everyone can infect everyone else in the same socio-physical compartment.

1. The emergence of a network-based framework has brought a fundamental change, offering a new level of predictability.
2. Today epidemic prediction is one of the most active applications of network science being used to foresee the spread of influenza or to contain Ebola.

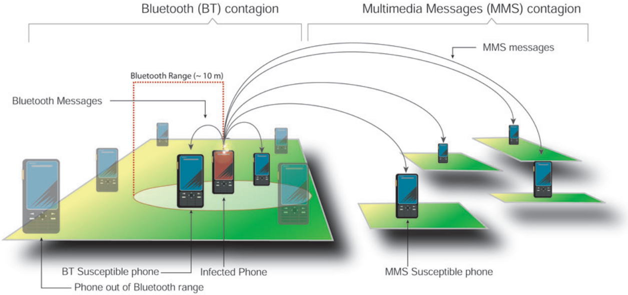
The emergence of a network-based framework has brought a fundamental change, offering a new level of predictability. Today epidemic prediction is one of the most active applications of network science being used to foresee the spread of influenza or to contain Ebola.

Social impacts

Beyond Epidemiology

In January 2010 network science tools have predicted the conditions necessary for the emergence of viruses spreading through mobile phones. The first major mobile epidemic outbreak that started in the fall of 2010 in China, infecting over 300,000 phones each day, closely followed the predicted scenario.

Mobile phone viruses



Scientific Impact

Nowhere is the impact of network science more evident than in the scientific community.

Complex Systems and Networks

The most prominent scientific journals, from Nature to Science, Cell and PNAS, have devoted reviews and editorials addressing the impact of networks on various topics, from biology to social sciences. For example, Science has published a special issue on networks, marking the ten-year anniversary of the discovery of scale-free networks.

Complex Systems and Networks

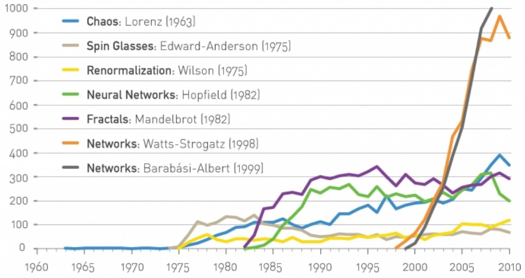
Figure : Special issue of Science magazine devoted to networks, published on July 24, 2009, on the 10th anniversary of the 1999 discovery of scale-free networks.

Scientific Impact

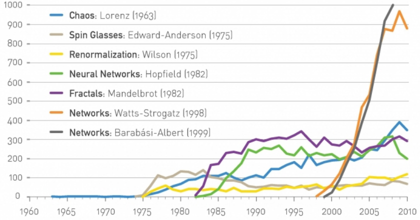
1. During the past decade each year about a dozen international conferences, workshops, summer and winter schools have focused on network science.
2. A highly successful network science conference series, called Net- Sci, attracts the fields practitioners since 2005.
3. Several general-interest books have made bestseller lists in many countries, bringing network science to the general public.
4. Most major universities offer network science courses, attracting a diverse student body, and in 2014 Northeastern University in Boston and the Central European University in Budapest have launched PhD programs in network science.

Scientific Impact

To see the impact of networks on the scientific community it is useful to inspect the citation patterns of the most cited papers in the area of complex systems. Each of these papers are citation classics, reporting classic discoveries like the butterfly effect, renormalisation group, spin glasses, fractals and neural networks.



Scientific Impact



To see how the interest in network science compares to the impact of these foundational papers their citation patterns is compared to the citations of the two most cited network science papers: the 1998 paper on small-world phenomena (Watt and Strogatz) and the 1999 Science paper reporting the discovery of scale-free networks (Barabasi and Albert). As one can see, the rapid rise of yearly citations to these two papers is without precedent in the area of complex systems.

Exercises/tutorial

1. Networks Everywhere
   List three different real networks and state the nodes and links for each of them.
2. Your Interest Tell us of the network you are personally most interested in. Address the following questions:
   1. What are its nodes and links?
   2. How large is it?
   3. Can be mapped out?
   4. Why do you care about it?
3. Impact
   In your view what would be the area where network science could have the biggest impact in the next decade? Explain your answer.