Rensselaer

why not change the world?®

Intro to Network Science

Ahmed Eleish

Data Analytics ITWS-4600/6600 CSCI-4960 MGMT 4600/6600

April 7th 2026

Tetherless World Constellation
Rensselaer Polytechnic Institute

Network

: an interconnected or interrelated chain, group, or system

: a usually informally interconnected group or association of persons (such as friends or professional colleagues)

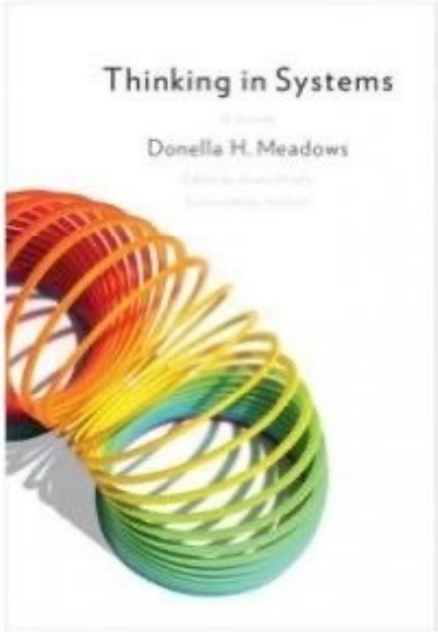
https://www.merriam-webster.com/dictionary/network

Contents

Background

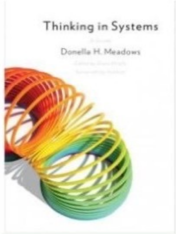
Systems

- Consists of primarily three things (Meadows)
  - Elements
  - Interconnections
  - Function/ Purpose



https://www.amazon.com/Thinking-Systems-Donella-H-Meadows/dp/1603580557

System:



–**A system has a purpose** - such as to distribute water to plant life, bouncing a communications signal around the country to consumers, or producing information for people to use in conducting business on in an organization.

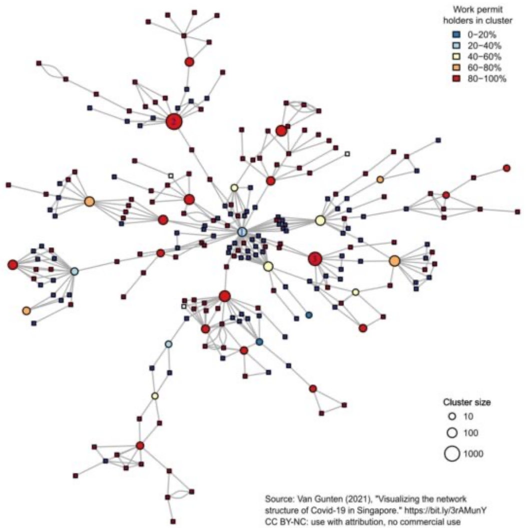
–**A system is a grouping of two or more components** which are held together through some common and cohesive bond. The bond may be water as in the irrigation system, a microwave signal as used in communications, or, as we will see, data in an information system.

–**A system operates routinely** and, as such, it is predictable in terms of how it works and what it will produce.

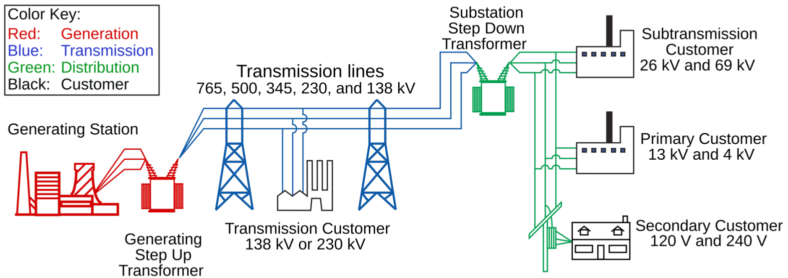
Book is available on Amazon:
https://www.amazon.com/Thinking-Systems-Donella-H-Meadows/dp/1603580557

Information embedded in Interconnectedness

- Networks model systems naturally



Covid-19 spread in Singapore



Simple diagram of electricity grids in North America

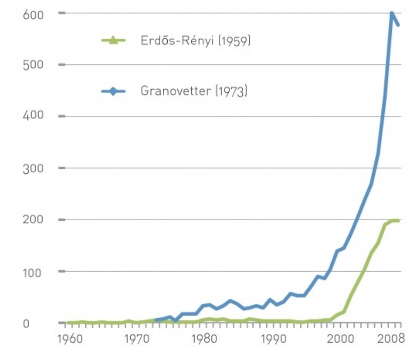
Networks are ubiquitous

- Science: gene, protein, fossil, mineral networks
- Infrastructure: telecommunication, electricity, transportation networks
- People: social, collaboration, disease transmission networks
- Digital: knowledge graphs, neural networks, virtual networks

2 Seminal papers in network science

- P. Erdős and A. Rényi. On random graphs. Publicationes Mathematicae, 6: 290, 1959.

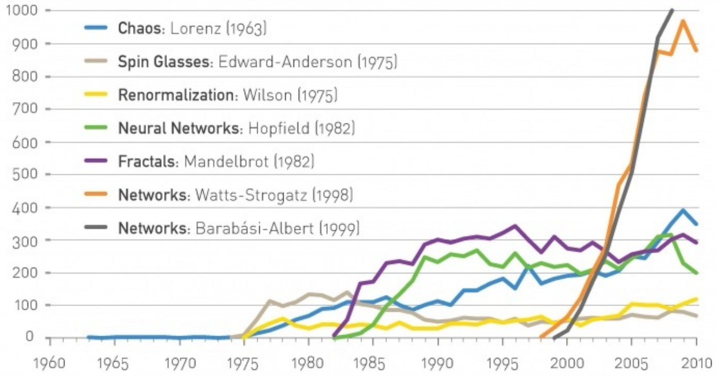
- M. S. Granovetter. The strength of weak ties. American Journal of Sociology 78: 1360, 1973.

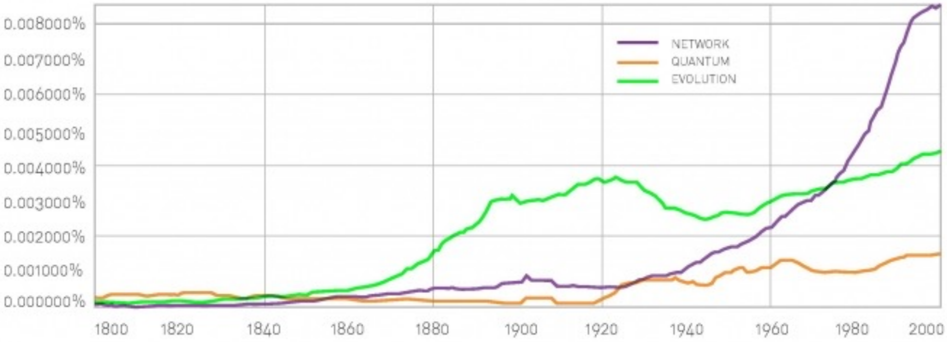


Network Science

- Interdisciplinary field taking a data-driven approach to understanding systems utilizing quantitative and computational methods.
- Despite its old foundations, Network Science has mostly emerged in the 21st century, highly popularized by social network analysis.
- Increasingly being used to learn about and teach complex systems in many domains.

The rising popularity of networks and network science from as seen in bibliometrics

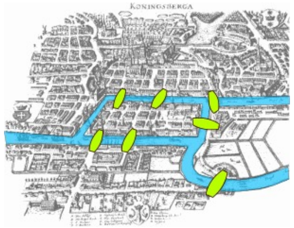




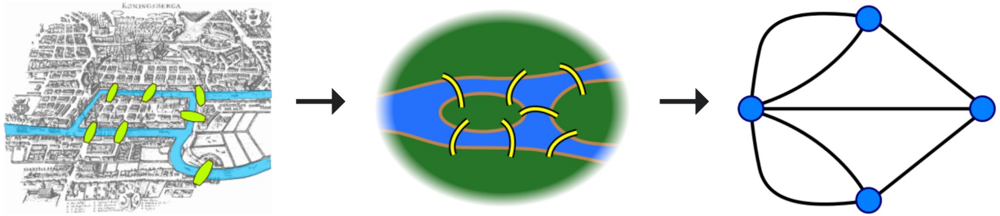
Graph Theory

The Bridges of Königsberg

Graph Theory was introduced by the Swiss mathematician Leonhard Euler during the 18th century with his famous problem known as the Seven Bridges of Königsberg.



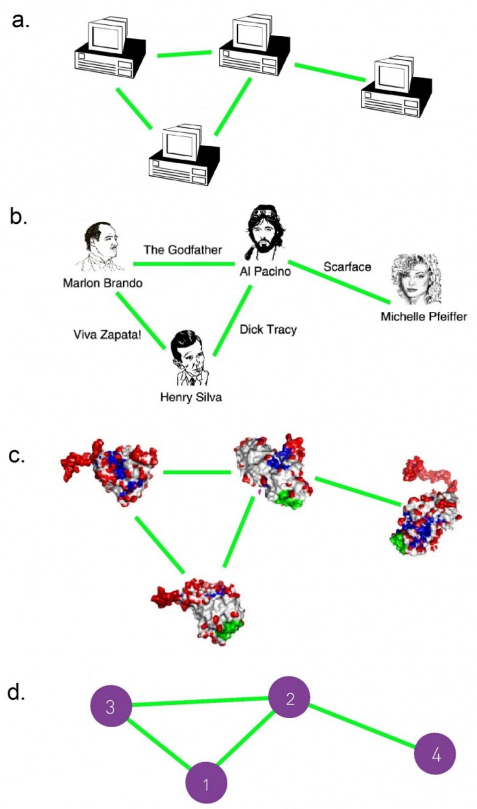
Euler used this graph to answer the question: Can you walk through the city and cross each bridge only once?



Images/Resource Credits: https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

Graphs of Networks

- Graphs are the mathematical representation of networks
- Networks consist of ***nodes*** and ***links***
- Graphs consist of ***vertices*** and ***edges***
- The grounding in graph theory provides network science/analysis with a quantitative basis for learning from the patters and structure embedded in networks



Basics of Network Analysis

Definitions

- Network: a collection of nodes representing entities, connected by links representing relationships between the entities.
- Node: the basic unit of a network, represents an entity within a system being studied. Nodes can be annotated by changing their shape, color, size to represent attributes of the entity.
- Link: a connection between 2 nodes, representing some relationship between the entities represented by the nodes. Links may have a direction as well as a weight (quantity) representing for example the strength of the connection.

Basics

- Networks typically consist of one type of node (unipartite). Although networks with 2 (bipartite), 3 (tripartite) or more (multipartite?) can be constructed, analyzing them becomes more difficult.

- Similarly with edges, networks with multiple types of edges are possible but less useful for direct analysis.

- Directed / undirected networks: networks may be undirected (no natural directions on edges), e.g. co-authorship network or they may be directed, e.g. webpage link network.

- Weighted / unweighted networks: in unweighted networks all edges have the same weight while in weighted networks, edges carry a weight that can represent a quality of the connection being represented.

Quantifying network structure

- Node degree
- The degree of a node, usually uses the symbol $\boldsymbol{k}$, in an undirected network is equal to the number of links connecting it to other nodes.
- In a directed network, a node has 2 degree values, $\boldsymbol{k_{in}}$ for incoming links and $\boldsymbol{k_{out}}$ for outgoing links.

- Average degree of a network
- It is useful to calculate the average degree of a network.

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i$$

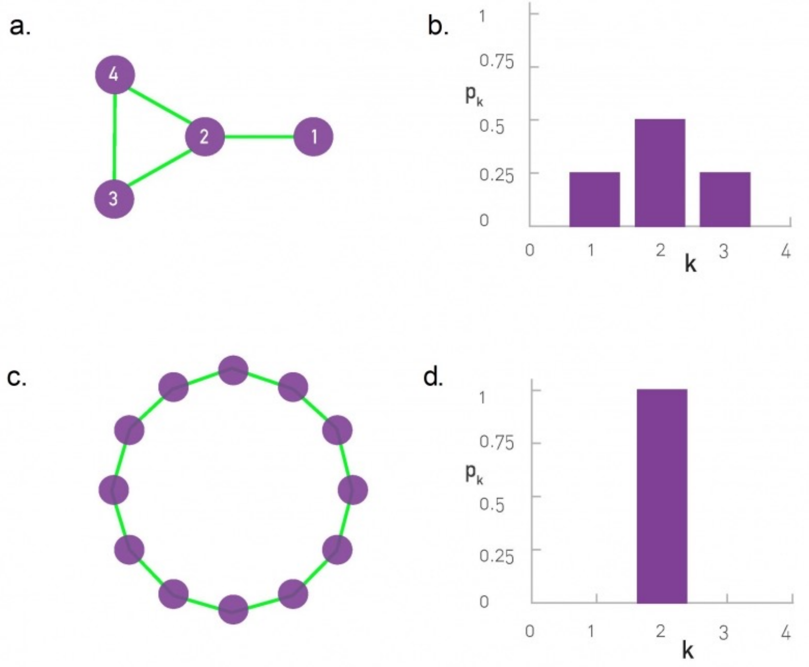
- This is similar to the branching factor of a tree.

Quantifying network structure

- Degree distribution of a network

- The degree distribution of a network demonstrates the probability of obtaining a node of a particular degree

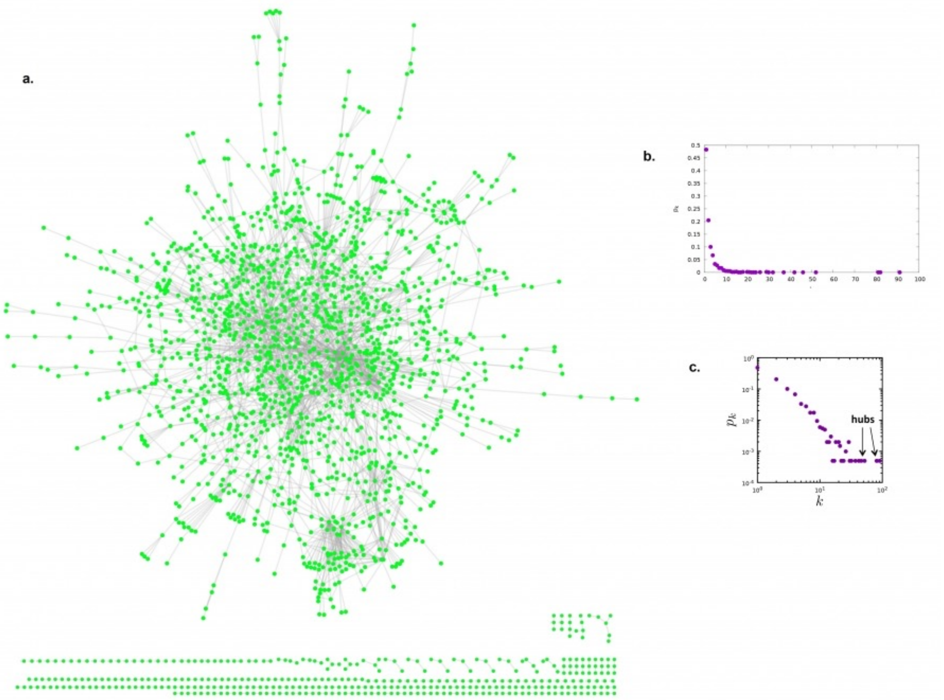
- Degree distributions represent a way to quantify branching structure in a network and compare across different networks.



Quantifying network structure

- Degree distribution of a network protein interaction network.

- Most nodes have a low degree value, a smaller number have a higher values and a few have a high degree value.



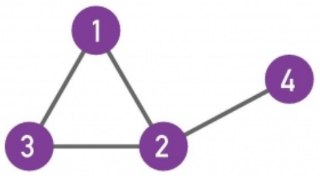
Representing networks mathematically

$$A_{ij} = \begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{matrix}$$

- **Adjacency matrix**

- A square, symmetric matrix with node numbers on the rows and columns, where a cell contains a value of 1 if the 2 nodes represented by the cell are connected by a link, and a value of 0 if the nodes are not connected.

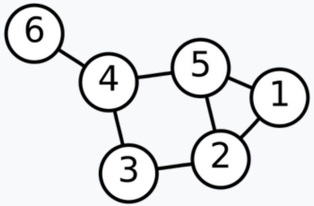
b. **Undirected network**



$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

Representing networks mathematically

- Other representations can be constructed from the same network

Labelled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$

https://en.wikipedia.org/wiki/Laplacian_matrix

To obtain the Laplacian:
L = D - A

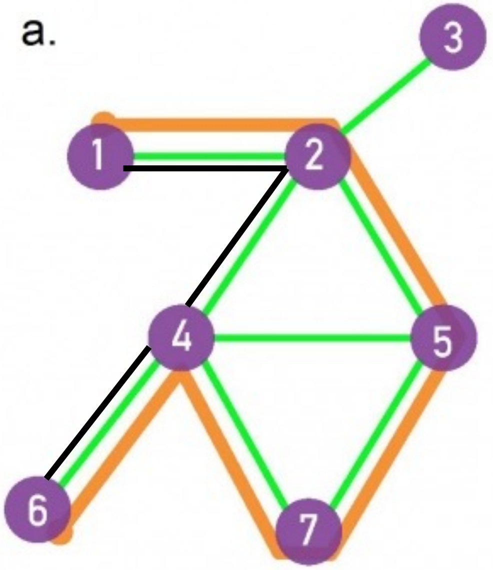
Quantifying network structure

- Paths and distances in networks

- Distance in networks is measured by the number of connections (hops) between nodes

- Typically, the shortest paths between nodes are examined to learn about the network structure.

- Shortest paths and distance can be calculated from the adjacency matrix.



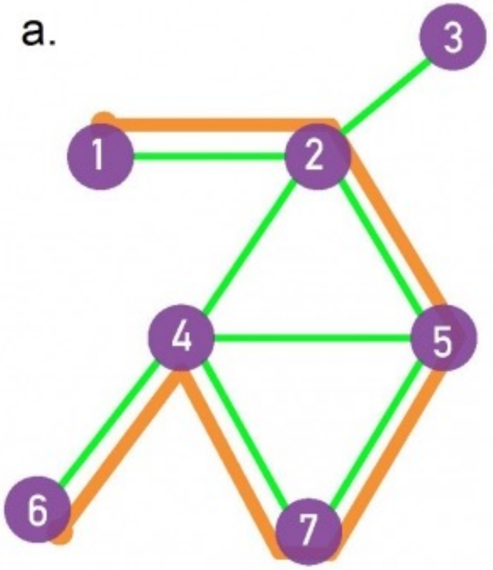
The orange path above between nodes 1 and 6 (1-2-5-7-5-6) has length 5 while the black path (1-2-4-6) has length 3

Quantifying network structure

- Network level measures

- Diameter: the maximum shortest path in the network / the largest distance between 2 nodes in the network.

- Density: ratio between the number of edges in a network and the maximum number of possible edges



Diameter = 3

Density = 8/21 = 0.38

Quantifying network structure

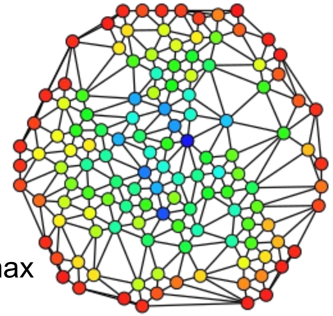
- Node level measures

- Node centrality: values describing the position of the node relative to other nodes in the network.

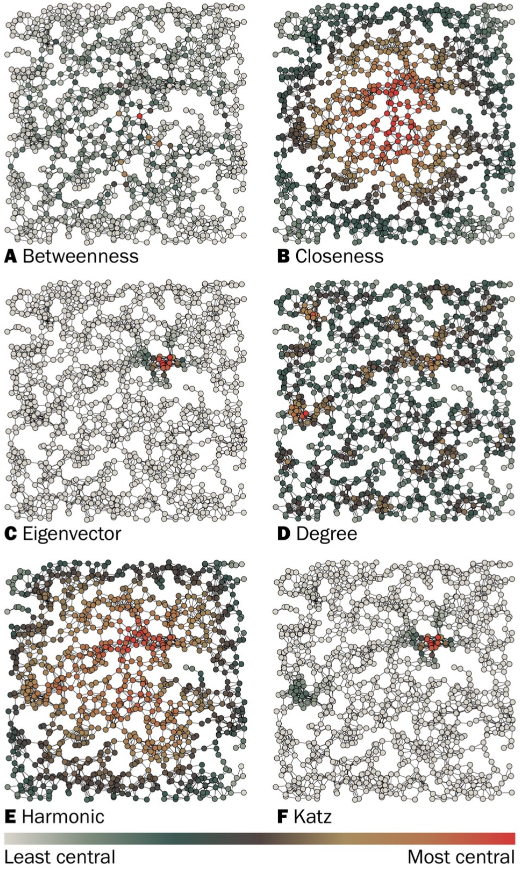
- There are many ways to quantify centrality, e.g. betweenness centrality quantities the number of shortest paths going through a node.

- More central nodes are considered **"hubs"** in the network.

Betweenness Centrality



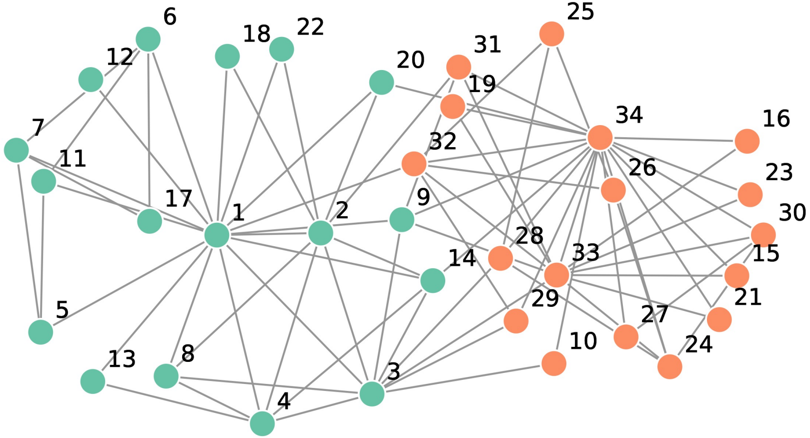
red = 0
blue = max



Examples from Research

Zachary's karate club

- Social network of a university karate club
- Nodes represent club members including the administrator “John A” and instructor “Mr. Hi” (pseudonyms)
- Links represent members interacting outside the club
- A conflict between the instructor and administrator resulted in the club splitting



Karate club network colored by membership in post-split faction

Quantifying ecological impacts of mass extinctions with network analysis of fossil communities

A. D. Muscente ✉, Anirudh Prabhu, Hao Zhong, Ahmed Eleish, Michael B. Meyer, Peter Fox, Robert M. Hazen, and Andrew H. Knoll ✉ -4 Authors Info & Affiliations

Contributed by Andrew H. Knoll, March 20, 2018 (sent for review November 15, 2017; reviewed by Mary L. Droser and Seth Finnegan)

April 23, 2018 | 115 (20) 5217-5222 | https://doi.org/10.1073/pnas.1719976115

Analysis of Fossil Networks

- Mass extinctions documented by the fossil record provide critical benchmarks for assessing changes through time in biodiversity and ecology.
- Efforts to compare biotic crises of the past and present, however, encounter difficulty because taxonomic and ecological changes are decoupled.
- Network analysis of fossil co-occurrence data enables us to identify nonrandom associations of interrelated paleo-communities.

Dataset

Supporting Information (SI) Dataset for "Quantifying ecological impacts of mass extinctions with network analysis of fossil communities" paper by Muscente et al. in PNAS

WORKBOOK PAGE	CONTENTS
o_taxa	List of animal orders that were included in (and excluded from) the "o" networks with non-weighted (binary) and weighted edges (revised taxonomy, module assignment, and data accessed from the Paleobiology Database)
o-n_diversity_major	Diversity through time data for the modules in the "o-n" network with non-weighted (binary) edges (data accessed from the Paleobiology Database) calculated using the "major method"
o-n_diversity_contain	Diversity through time data for the modules in the "o-n" network with non-weighted (binary) edges (data accessed from the Paleobiology Database) calculated using the "contain method"
o-w_diversity_major	Diversity through time data for the modules in the "o-w" network with weighted edges (data accessed from the Paleobiology Database) calculated using the "major method"
o-w_diversity_contain	Diversity through time data for the modules in the "o-w" network with weighted edges (data accessed from the Paleobiology Database) calculated using the "contain method"
f_taxa	List of animal families that were included in (and excluded from) the "f" networks with non-weighted (binary) and weighted edges (revised taxonomy, module assignment, and data accessed from the Paleobiology Database)
f-n_diversity_major	Diversity through time data for the modules in the "f-n" network with non-weighted (binary) edges (data accessed from the Paleobiology Database) calculated using the "major method"
f-n_diversity_contain	Diversity through time data for the modules in the "f-n" network with non-weighted (binary) edges (data accessed from the Paleobiology Database) calculated using the "contain method"
f-w_diversity_major	Diversity through time data for the modules in the "f-w" network with weighted edges (data accessed from the Paleobiology Database) calculated using the "major method"
f-w_diversity_contain	Diversity through time data for the modules in the "f-w" network with weighted edges (data accessed from the Paleobiology Database) calculated using the "contain method"
g_taxa	List of animal genera that were included in the "g1," "g2," and "g3" networks with non-weighted (binary) and weighted edges (revised taxonomy, module assignment, and data accessed from the Paleobiology Database)
g1-n_proportions	Proportions of modules in genus-level (g1-n) network of marine animal genera (data used in Fig. 3)
g1-n_diversity_major	Diversity through time data for the modules in the "g1-n" network with non-weighted (binary) edges (data accessed from the Paleobiology Database) calculated using the "major method"
g1-n_diversity_contain	Diversity through time data for the modules in the "g1-n" network with non-weighted (binary) edges (data accessed from the Paleobiology Database) calculated using the "contain method"
g1-w_diversity_major	Diversity through time data for the modules in the "g1-w" network with weighted edges (data accessed from the Paleobiology Database) calculated using the "major method"
g1-w_diversity_contain	Diversity through time data for the modules in the "g1-w" network with weighted edges (data accessed from the Paleobiology Database) calculated using the "contain method"
g2-n_diversity_major	Diversity through time data for the modules in the "g2-n" network with non-weighted (binary) edges (data accessed from the Paleobiology Database) calculated using the "major method"
g2-w_diversity_major	Diversity through time data for the modules in the "g2-w" network with weighted edges (data accessed from the Paleobiology Database) calculated using the "major method"
g3-n_diversity_major	Diversity through time data for the modules in the "g3-n" network with non-weighted (binary) edges (data accessed from the Paleobiology Database) calculated using the "major method"
g3-w_diversity_major	Diversity through time data for the modules in the "g3-w" network with weighted edges (data accessed from the Paleobiology Database) calculated using the "major method"
non-weighted_swings_major	Total swing calculations based on diversity through time data ("major method") for modules in "o-n," "f-n," and "g1-n" networks with non-weighted (binary) edges
non-weighted_swings_contain	Total swing calculations based on diversity through time data ("contain method") for modules in "o-n," "f-n," and "g1-n" networks with non-weighted (binary) edges
weighted_swings_major	Total swing calculations based on diversity through time data ("major method") for modules in "o-w," "f-w," and "g1-w" networks with weighted edges
weighted_swings_contain	Total swing calculations based on diversity through time data ("contain method") for modules in "o-w," "f-w," and "g1-w" networks with weighted edges
paleoenvironments	Counts and percentages of the occurrecnes of the modules in this study from the various paleoenvironments

Community Detection in Fossil Networks

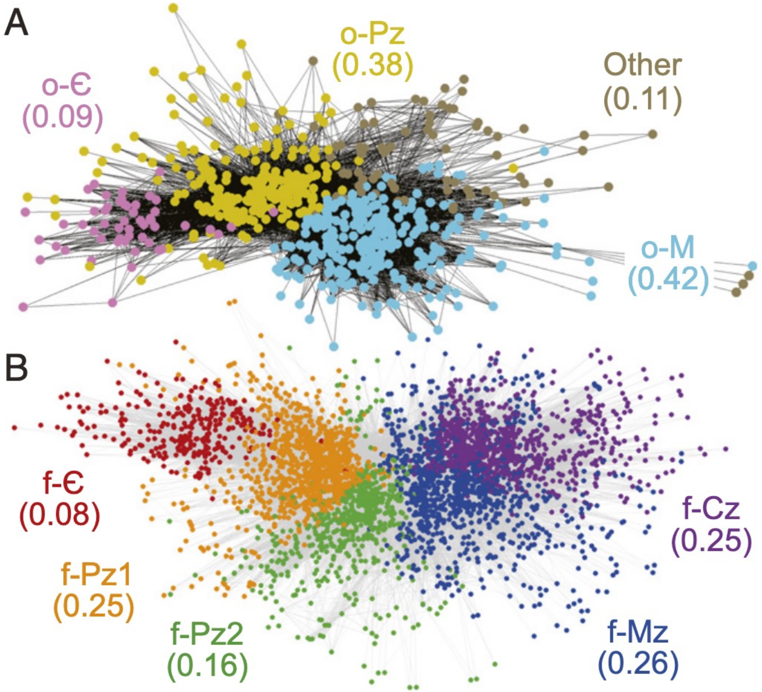


Fig. 1. Graphs of nonweighted networks partitioned with the Louvain algorithm. (*A*) The *o–n* network of marine animal orders (*n* = 621), which fall into three main modules, resembling the Cambrian, Paleozoic, and Modern evolutionary faunas (1). (*B*) The *f–n* network of marine animal families (*n* = 3,729), which fall into five modules, similar to the main clusters in the corresponding generic-level (*g1–n* and *g2–n*) networks. Values in parentheses indicate the modules' proportions of their respective networks. See figures in *SI Appendix* for additional graphs (*SI Appendix*, Figs. S4–S8).

Diversity vs Geologic Age

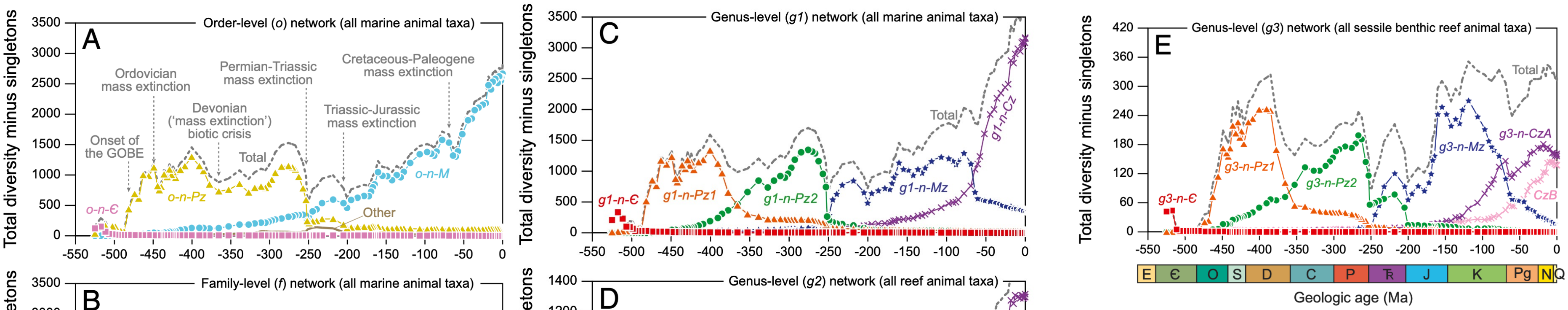


Fig. 2. Plots show diversity versus geologic age for modules in the non-weighted networks. The value of each bin equals the total number of boundary crossing genera, i.e., total diversity minus singletons (4). (*A*) The *o–n* network of marine animal orders (Fig. 1*A* and *SI Appendix*, Fig. S4*A*). (*B*) The *f–n* network of marine animal families (Fig. 1*B* and *SI Appendix*, Fig. S5*A*). (*C*) The *g1–n* network of marine animal genera (n= 26,525; *SI Appendix*, Fig. S6*A*). (*D*) The *g2–n* network of reef-occurring animal genera (*n* = 7,621; *SI Appendix*, Fig. S7*A*). (*E*) The *g3–n* network of reef-building animal genera (*n* = 2,793; *SI Appendix*, Fig. S8*A*).

Diversity Swing vs Geologic Age

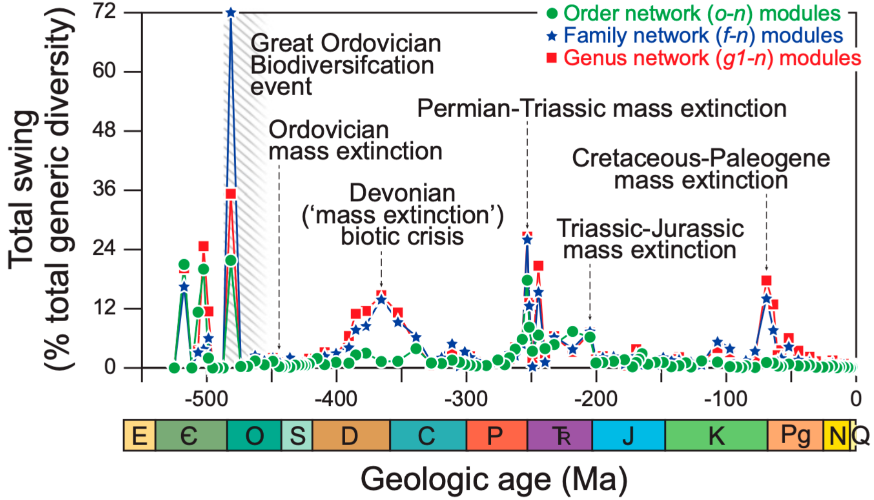


Fig. 4. Ecological severities of biotic crises and critical transitions. Plot shows the total swing in diversity of each Phanerozoic (post-Fortunian) geologic stage for the nonweighted (*o–n*, *f–n*, and *g1–n*) networks. A swing is a relative change in representation (i.e., growth of one module and reduction of another) across an interval, and is expressed in terms of percent total diversity minus singletons (data downloaded from the PBDB using the contain method of time binning). To calculate total swing for a geologic stage, the absolute difference in percent total diversity between the upper and lower boundaries was determined for each module. The sum of these values was then determined and divided by 2. Results did not significantly vary with binning approach (*SI Appendix*, Fig. S13).

 Open Access Published by De Gruyter July 31, 2017

Network analysis of mineralogical systems

Shaunna M. Morrison, Chao Liu, Ahmed Eleish, Anirudh Prabhu, Congrui Li, Jolyon Ralph, Robert T. Downs, Joshua J. Golden, Peter Fox, Daniel R. Hummer, Michael B. Meyer and Robert M. Hazen

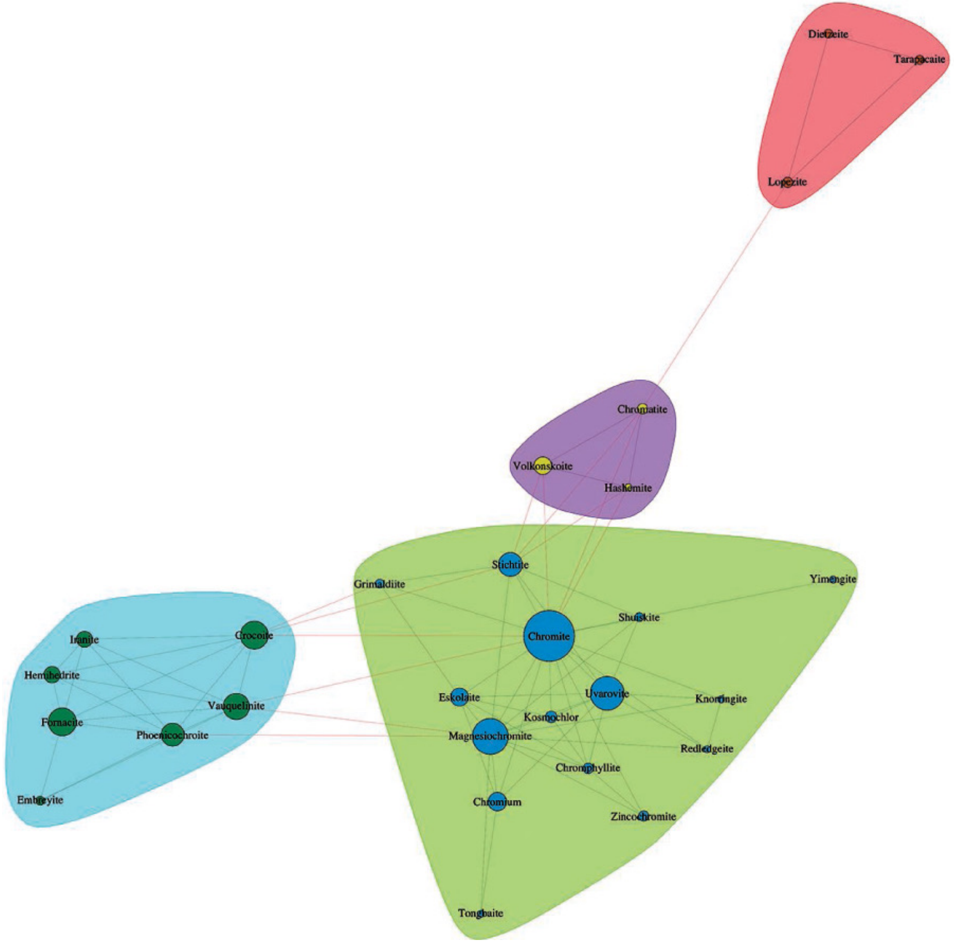
From the journal American Mineralogist
https://doi.org/10.2138/am-2017-6104CCBYNCND

Analysis of Mineral Networks

- This approach provides a dynamic visualization platform for higher-dimensional analysis of phase relationships, because topologies of equilibrium phase assemblages and pathways of mineral reaction series are embedded within the networks.
- Mineral networks also facilitate quantitative comparison of lithologies from different planets and moons, the analysis of coexistence patterns simultaneously among hundreds of mineral species and their localities, the exploration of varied paragenetic modes of mineral groups, and investigation of changing patterns of mineral occurrence through deep time.
- Mineral network analysis, furthermore, represents an effective visual approach to teaching and learning in mineralogy and petrology.

Mineral Networks (Chromium)

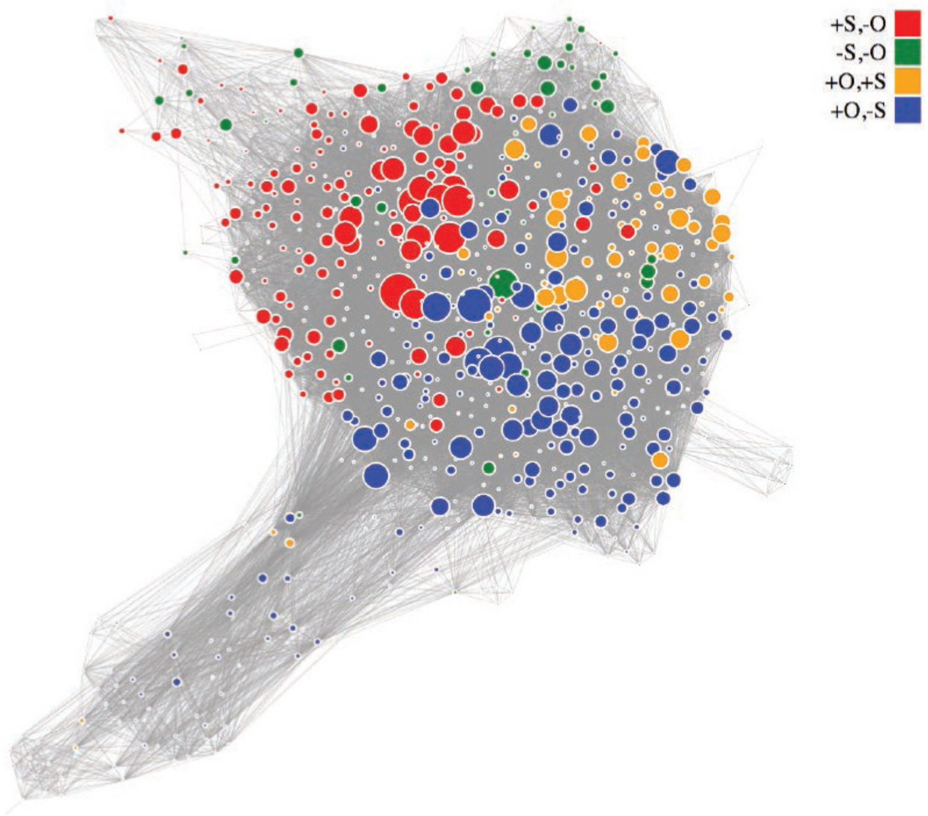
- Community Detection of network of 30 common chromium-bearing minerals reveals segregation into four groups.
- The central cluster (group 1) includes 17 Cr3+ species formed through igneous, metamorphic, or hydrothermal processes.
- The left-hand cluster (group 2) includes seven Cr6+ species formed through hydrothermal alteration.
- The two smaller clusters (groups 3 and 4) include chromate minerals precipitated in soils and desert environments.



Mineral Networks (Copper)

- 664 copper minerals sorted by composition

- *~ Copper bearing minerals with similar compositions tend to co-exist*



Mineral Network Metrics

- Compare mineral systems across lithologies, compositions, planets, etc.

Mineral system	Density	Centralization	Transitivity	Diameter	Mean distance
Igneous minerals	0.64	0.34	0.77	2	1.36
Cr minerals	0.05	0.33	0.44	6	2.65
Cu minerals	0.12	0.68	0.48	4	1.93

More Mineral Networks

https://dtdi.carnegiescience.edu/node/4557

Summary

- Networks are useful for studying complex systems; Network Science is an increasingly popular interdisciplinary field of research applied in various domains.
- Network analysis is a suite of methods and tools to represent, construct, study, and learn from networks.
- Excellent resource: https://networksciencebook.com/ by Albert-László Barabási

Thanks!