Binghamton University

EngiNet™

State University of New York

WARNING

Contact Information:

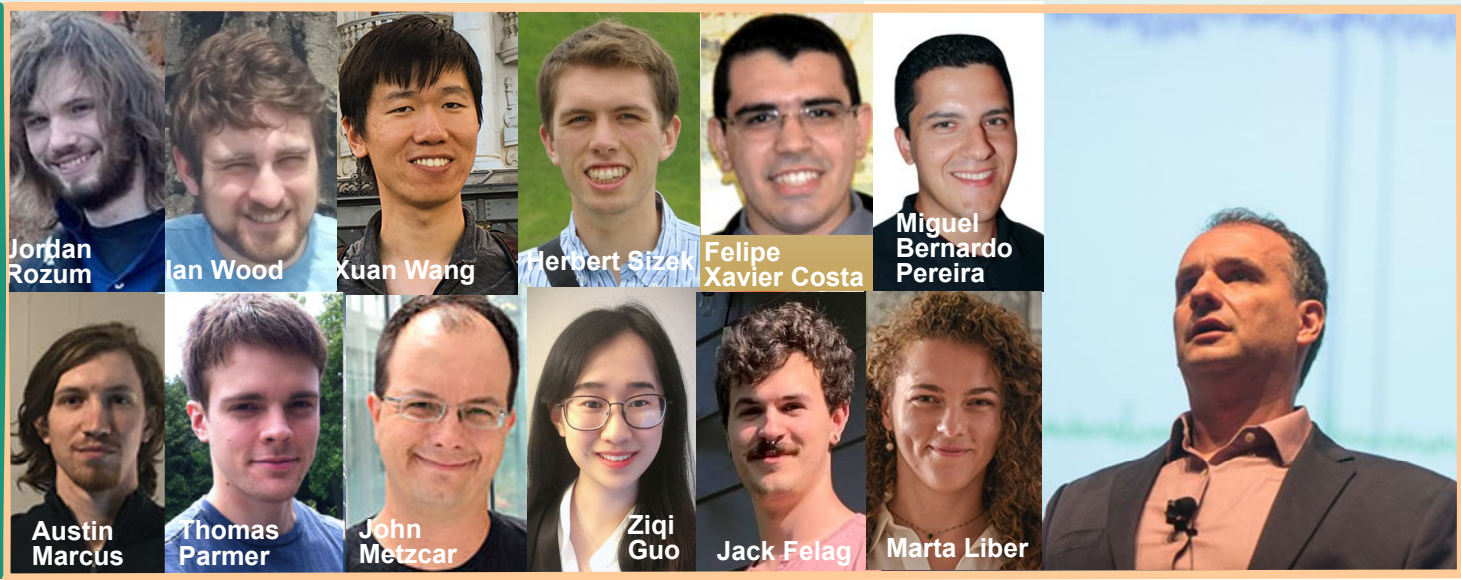
EngiNet Office Staff:
Janice Kinzer
Email: enginet@binghamton.edu
Phone: 1-800-478-0718 or 607-777-4965

Media Production Operator: **Sneha Rawat**

Instructor: Prof. **Luis M. Rocha**
Email: rocha@binghamton.edu
Phone: 607-777-5934

TA: **????**

interdisciplinary science



for understanding social and biomedical complexity

luis m. rocha

FULBRIGHT

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

cnrs

NSF

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

FUNDAÇÃO LUSO-AMERICANA

IARPA BE THE FUTURE

PERSISTENT

careerbuilder

NIH NLM

GULBENKIAN SCIENCE

rocha@binghamton.edu
informatics.indiana.edu/rocha

Complex adaptive systems and computational intelligence (casci lab)

Resources

- web page
  - casci.binghamton.edu/academics/ssie501
- online class
  - binghamton.zoom.us/j/93351260610
- blog: sciber
  - sciber.blogspot.com
- Brightspace
  - brightspace.binghamton.edu/d2l/home/255004

SSIE-501/ISE-440 - Fall 2022

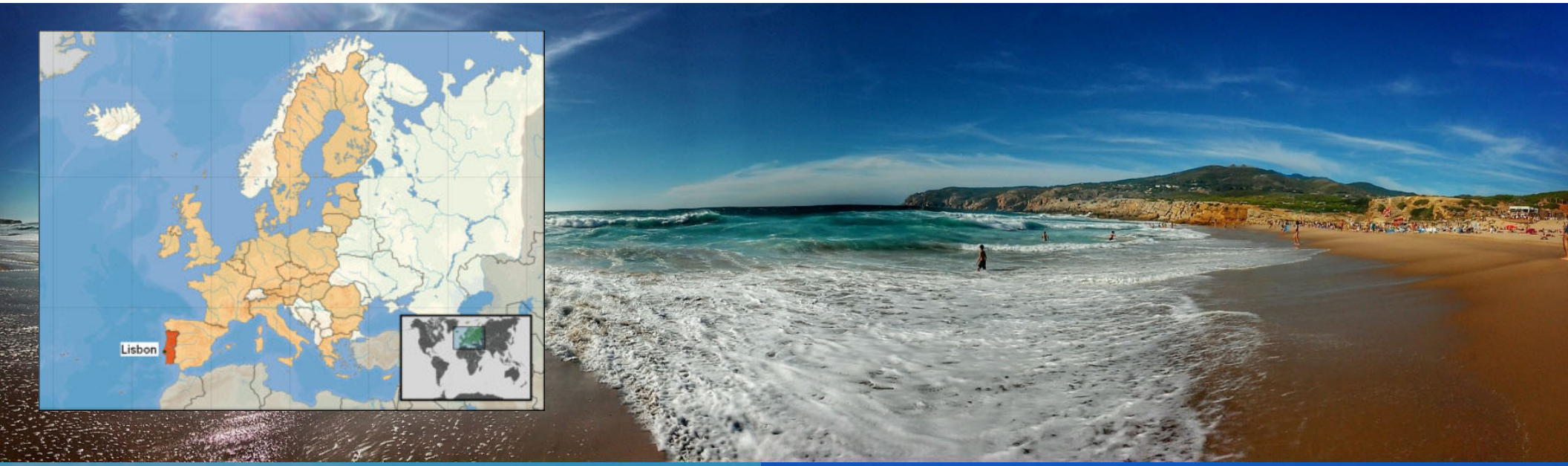
luis m. rocha

office hours:
Tuesdays 9:00- 11:30am
binghamton.zoom.us/my/luismrocha

office hours:
Tuesdays: 7:00-8:00pm????
binghamton.zoom.us/my/

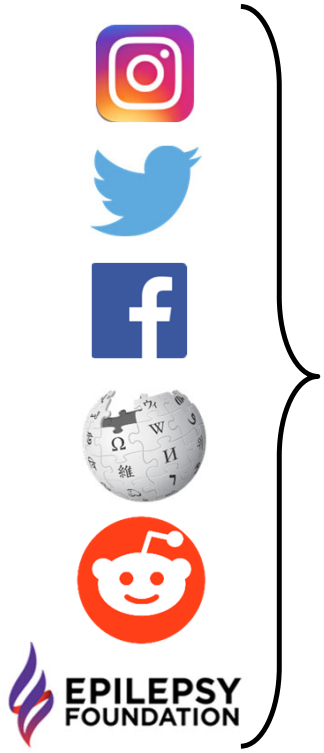
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK


Lisbon


Luanda

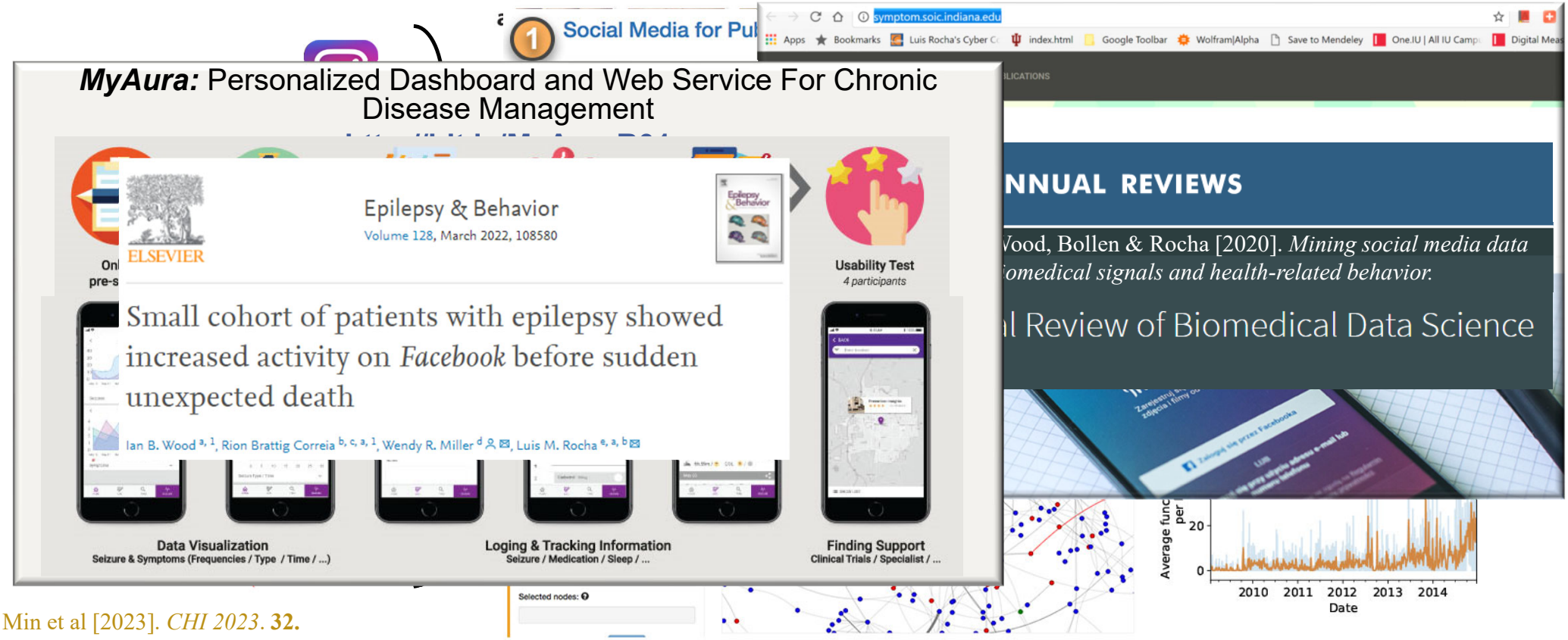
social media data pipelines for biomedicine



Min et al [2023]. *CHI 2023*. **32.**
Wood, Correia, Miller, &Rocha [2022]. *Epilepsy & Behavior*. **128**: 108580.
Correia, Wood, Bollen, & Rocha [2020]. *Annual Review of Biomedical Data Science*, 3:1.
Wood, Varela, Bollen, Rocha & Sá [2017]. *Scientific Reports. 7: 17973* .
Correia, Li & Rocha [2016]. *PSB*: **21**:492-503.
Ciampaglia, et al [2015]. *PloS ONE*. **10**(6): e0128193.

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

social media data pipelines for biomedicine



Min et al [2023]. *CHI 2023*. **32.**

Wood, Correia, Miller, &Rocha [2022]. *Epilepsy & Behavior*. **128**: 108580.

Correia, Wood, Bollen, & Rocha [2020]. *Annual Review of Biomedical Data Science*, 3:1.

Wood, Varela, Bollen, Rocha & Sá [2017]. *Scientific Reports. 7: 17973* .

Correia, Li & Rocha [2016]. *PSB*: **21**:492-503.

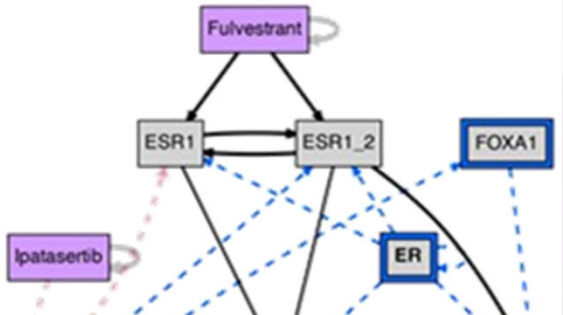
Ciampaglia, et al [2015]. *PloS ONE*. **10**(6): e0128193.

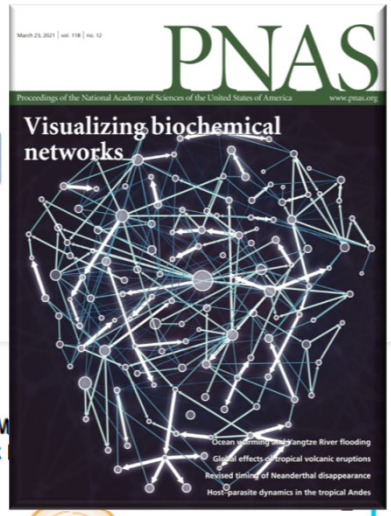
BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK

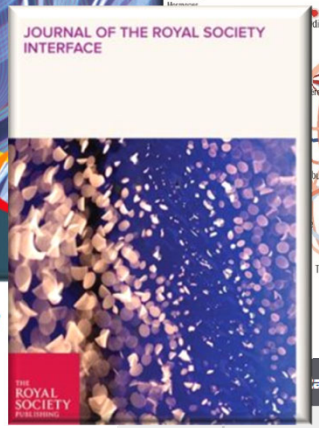
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

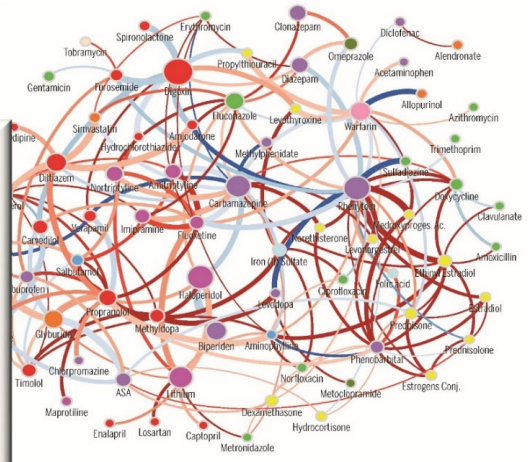
integrating and analyzing multiomic electronic health records with network science

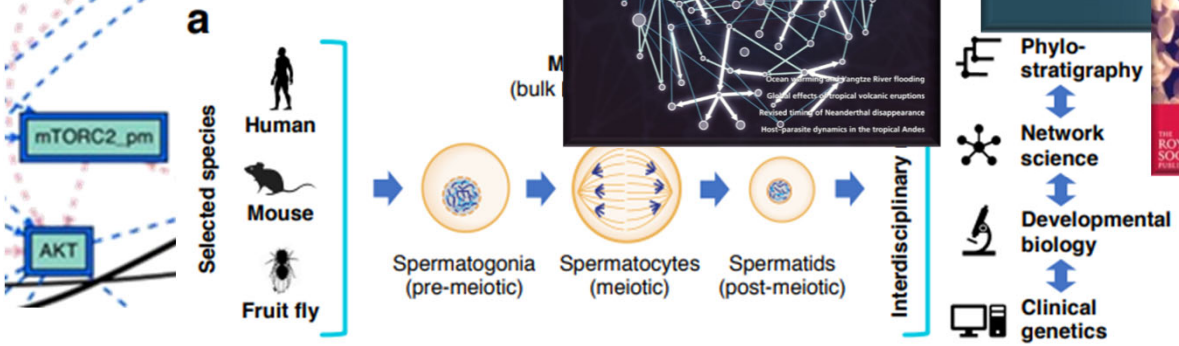
to predict comorbidity & drug interaction networks, disease factors & interventions

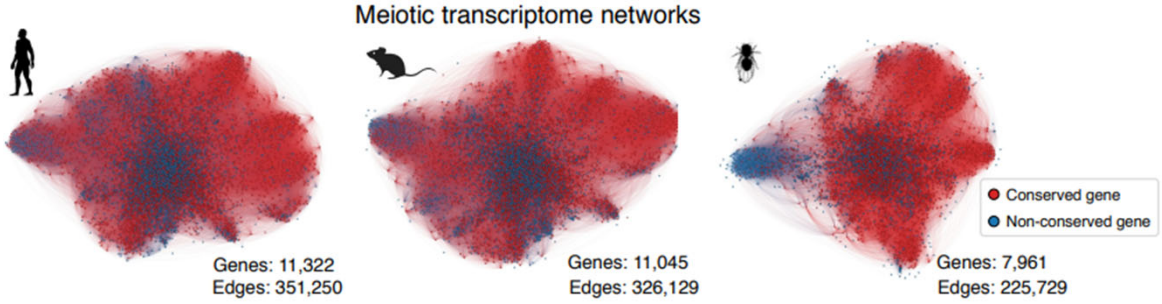












al medicine > articles > article

npj Digital Medicine

Correia, Araujo, Mattos & Rocha [2019]. **2**:*74*.

Article | Open Access | Published: 23 July 2019

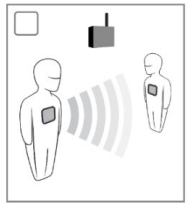
City-wide electronic health records reveal gender and age biases in administration of known drug–drug interactions

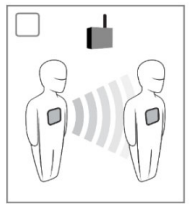
BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK

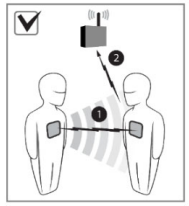
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

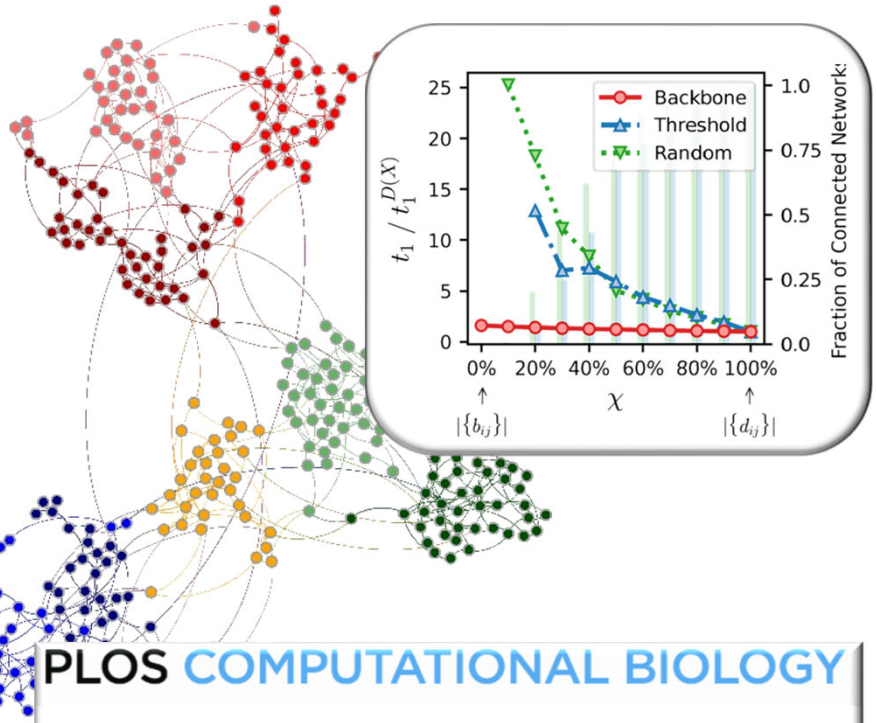
integrating and analyzing multilevel data sources with network science

to predict disease spread, information integration









PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

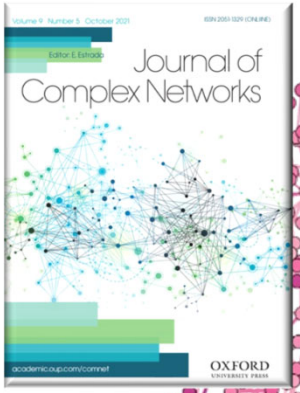
Contact networks have small metric backbones that maintain community structure and are primary transmission subgraphs

Rion Brattig Correia, Alain Barrat, Luis M. Rocha

frontiers in NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE

Multi-scale integration and predictability in resting state brain activity



Simas & Rocha [2015]. *Network Science*. doi:10.1017/nws.2015.11

Simas, Correia & Rocha [2021]. *J Complex Networks*. **9** (6), cnab021.

BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK

rocha@binghamton.edu

casci.binghamton.edu/academics/ssie501


E-TRASH
LIVE IN LISBON
TikTok
@djetrash
FEB 21 - 10 PM - NO COVER
FRIDAY, AUGUST 19: 1AM (BASEMENT)

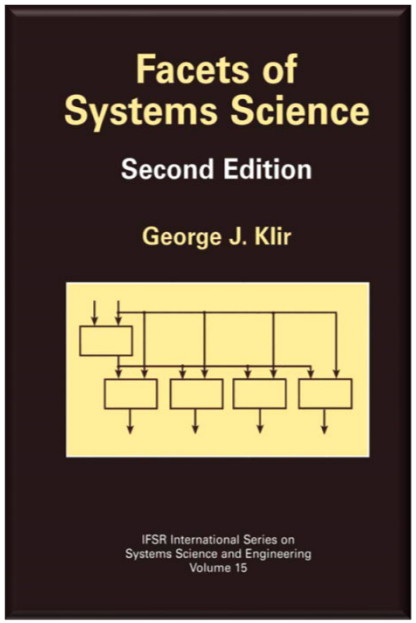
what about you?

- Background
- Interests
- Course expectations

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

course materials



- Lecture slides and notes
  - See course web page and brightspace
- Web links and general materials
  - Blog (**sciber.blogspot.com**) and brightspace
- Class Book
  - Klir, G.J. [2001]. *Facets of systems science*. Springer.
    - Available in electronic format for SUNY students.
- Various literature for discussion
  - Course web site and brightspace


George Klir

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

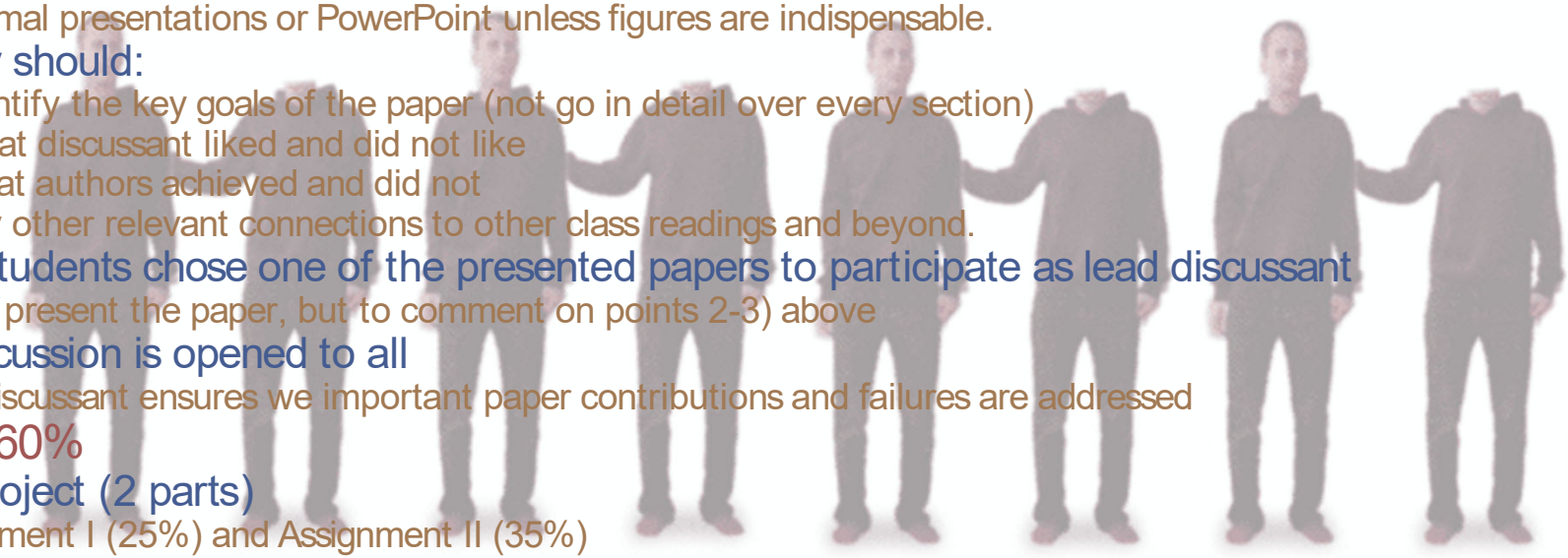
Overview and aims

- The course deals with the foundations of Systems Science, as well as current advances in Complex Networks and Systems which is the modern expression of this interdisciplinary field.
- Aims
  - Introduce and discuss the history, methodology and impact of complex systems science.
  - key literature, recent advances, and computational techniques in the field.
  - study concepts such as
    - Information, General Systems Theory, Networks, Modeling, Multi-Level Complexity, as well as their impact on science and society.
  - The course will also attempt to define and understand what systems thinking can bring to science and society.

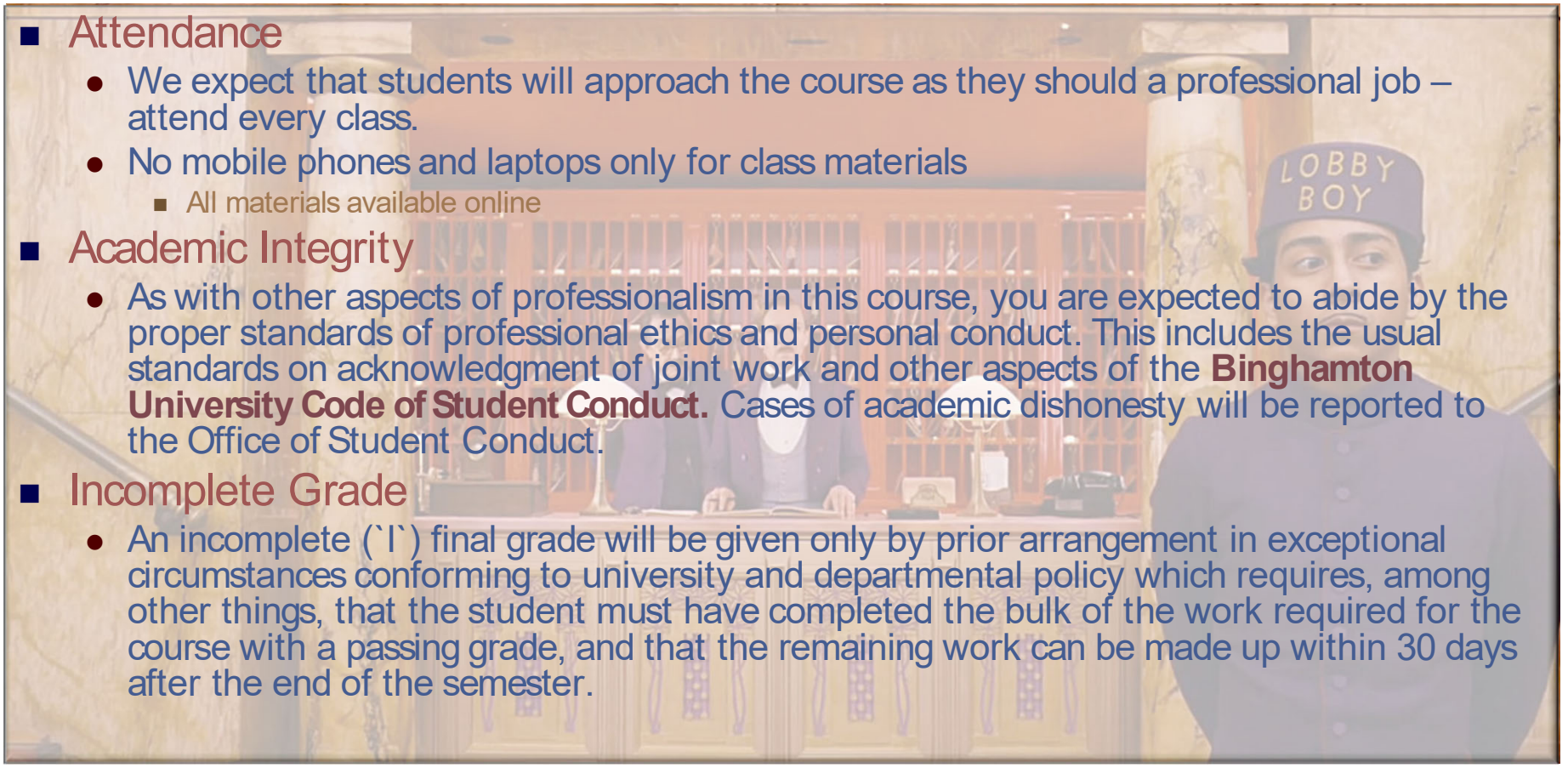
BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
casci.binghamton.edu/academics/ssie501

evaluation

- Participation: 20%.
  - class discussion, everybody reads and discusses every paper
  - engagement in class
- Paper Presentation and Discussion: 20%
  - SSIE501 students are assigned to papers as lead presenters and discussants
    - all students are supposed to read and participate in discussion of every paper.
  - Presenter prepares short summary of assigned paper (15 minutes)
    - no formal presentations or PowerPoint unless figures are indispensable.
  - Summary should:
    - 1) Identify the key goals of the paper (not go in detail over every section)
    - 2) What discussant liked and did not like
    - 3) What authors achieved and did not
    - 4) Any other relevant connections to other class readings and beyond.
  - ISE440 students chose one of the presented papers to participate as lead discussant
    - not to present the paper, but to comment on points 2-3) above
  - Class discussion is opened to all
    - lead discussant ensures we important paper contributions and failures are addressed
- Black Box: 60%
  - Group Project (2 parts)
    - Assignment I (25%) and Assignment II (35%)



but collegiality above all



- Attendance
  - We expect that students will approach the course as they should a professional job – attend every class.
  - No mobile phones and laptops only for class materials
    - All materials available online
- Academic Integrity
  - As with other aspects of professionalism in this course, you are expected to abide by the proper standards of professional ethics and personal conduct. This includes the usual standards on acknowledgment of joint work and other aspects of the **Binghamton University Code of Student Conduct.** Cases of academic dishonesty will be reported to the Office of Student Conduct.
- Incomplete Grade
  - An incomplete (`I`) final grade will be given only by prior arrangement in exceptional circumstances conforming to university and departmental policy which requires, among other things, that the student must have completed the bulk of the work required for the course with a passing grade, and that the remaining work can be made up within 30 days after the end of the semester.

for course

A+ **A** **A-**	**98%** **94** **90**	*Excellent Work.* Student performance demonstrates thorough knowledge of the course materials and exceeds course expectations by completing all requirements in a superior manner.
B+ **B** **B-**	**85** **80** **75**	*Very Good Work*. Student performance demonstrates above-average comprehension of the course materials and exceeds course expectations on all tasks as defined in the course syllabus.
C+ **C** **C-**	**70** **65** **60**	*Good Work*. Student performance meets designated course expectations and demonstrates understanding of the course materials at an acceptable level.
D+ **D** **D-**	**55** **50** **45**	*Marginal Work*. Student performance demonstrates incomplete understanding of course materials.
F	**Less than 45**	*Fail*.

key events coming up

- Paper Presentation: 20%
  - Present (501) and lead (501&440) the discussion of an article related to the class materials
    - Enginet students post/send video or join by Zoom synchronously
- Module 1: Cybernetics and the Information Turn
- Today
  - Borges, Jorge Luis. [1941]. *The Library of Babel*.
  - Borges, Jorge Luis. [1941]. *The Garden of Forking Paths* .
- Next classes
  - Presenter 1:
    - Heims, S.G. [1991]. *The Cybernetics Group*. MIT Press. Chapters: 1 - 2.
      - Optional: Chapters 11-12.
      - Optional: McCulloch, W. and W. Pitts [1943], "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics* **5**:115-133.
  - Presenter 2
    - Gleick, J. [2011]. *The Information: A History, a Theory, a Flood*. Random House. Chapter 8.
      - Optional: Prokopenko, Mikhail, Fabio Boschetti, and Alex J. Ryan. "An information theoretic primer on complexity, self-organization, and emergence." *Complexity* **15**.1 (2009): 11-28.
      - Kline, Ronald R [2015]. *The cybernetics moment, or, why we call our age the information age*. Johns Hopkins University Press. Chapters 1-2.
  - Presenter 3
    - Brenner, Sydney. [2012]. "History of Science. The Revolution in the Life Sciences". *Science* **338** (6113): 1427-8.
    - Brenner, Sydney. [2012]. "Turing centenary: Life's code script. *Nature* **482** (7386) (February 22): 461-461.
    - Cobb, Matthew. [2013]. "1953: When Genes Became 'Information'." *Cell* **153** (3): 503-506.
      - Optional: Searls, David B. [2010]. "The Roots of Bioinformatics". *PLoS Computational Biology* **6**(6): e1000809.
  - Presenter 4
    - Weaver, W. [1948]. "Science and Complexity". *American Scientist*, **36**(4): 536-44. Also available in Klir, G.J. [2001]. *Facets of systems Science*. Springer, pp: 533-540.
  - **Discussion by all**

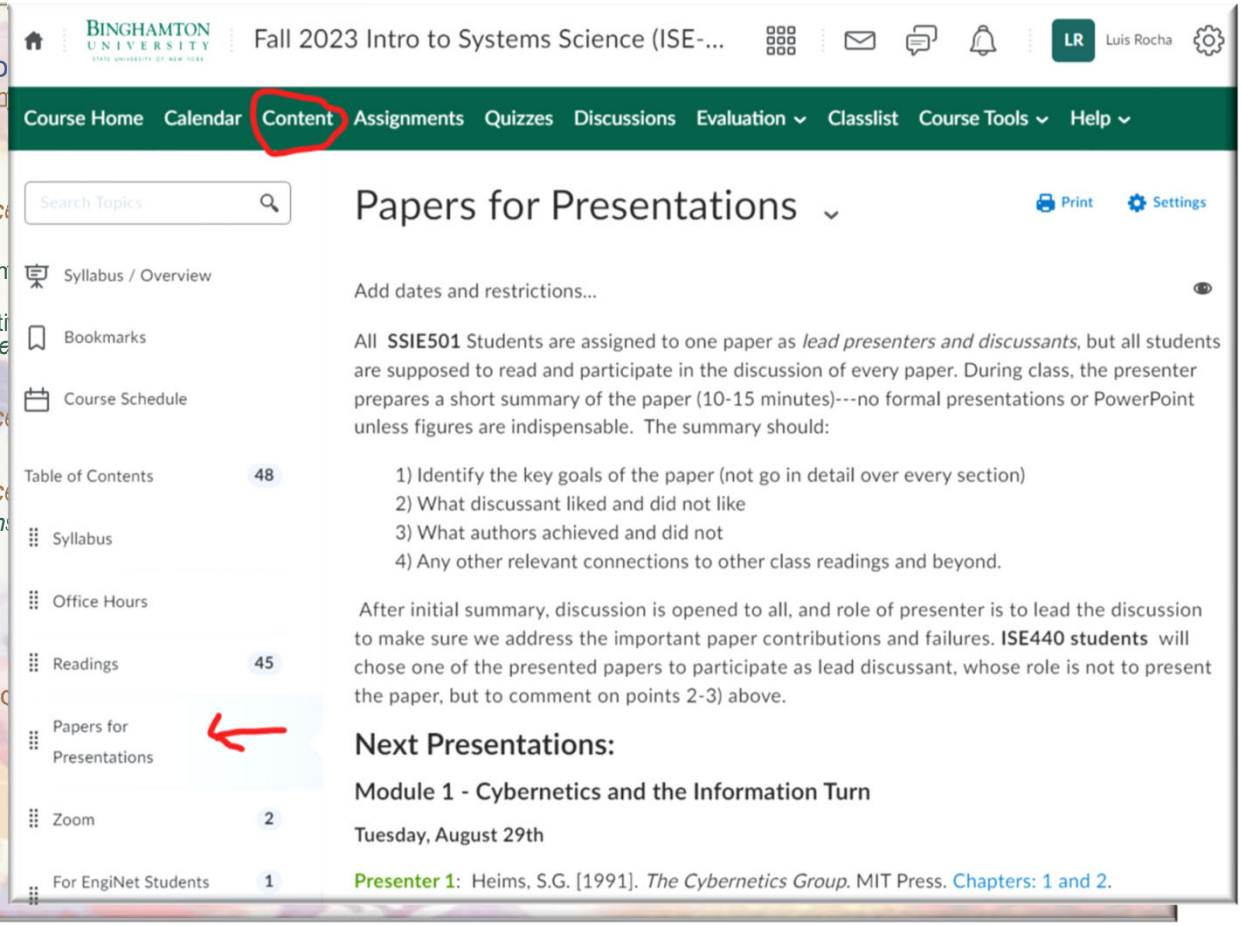
rocha@indiana.edu
informatics.indiana.edu/rocha

more upcoming readings (check brightspace)

- Paper Presentation: 20%
  - Present (501) and lead (501&440) the discussion of an article related to the class materials
    - Enginet students post/send video or join by Zoom synchronously
- Module 2: Systems Science
  - Presenter 5 :
    - Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapters 1 and 2.
      - Optional:
      - Rosen, R. [1986]. "Some comments on systems and system theory". *Int. J. of General Systems*, **13**: 1-3. Available in: Klir, G.J. [2001]. *Facets of systems Science*. Springer. pp: 241-243.
      - Wigner, E.P. [1960], "The unreasonable effectiveness of mathematics in the natural sciences". Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Comm. Pure Appl. Math*, **13**: 1-14.
  - Presenter 6:
    - Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapter 3.
  - Presenter 7:
    - Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapter 8.
      - Optional: Klir, G.J. [2001]. *Facets of systems Science*. Springer. Chapter 11
  - Presenter 8:
    - Schuster, P. (2016). The end of Moore's law: Living without an exponential increase in the efficiency of computational facilities. *Complexity*. **21**(S1): 6-9. DOI 10.1002/cplx.21824.
  - Presenter 9:
    - Von Foerster, H., P. M. Mora and L. W. Amiot [1960]. "Doomsday: Friday, November 13, AD 2026." Science **132**(3436):1291-5.
- Future Modules
  - See brightspace

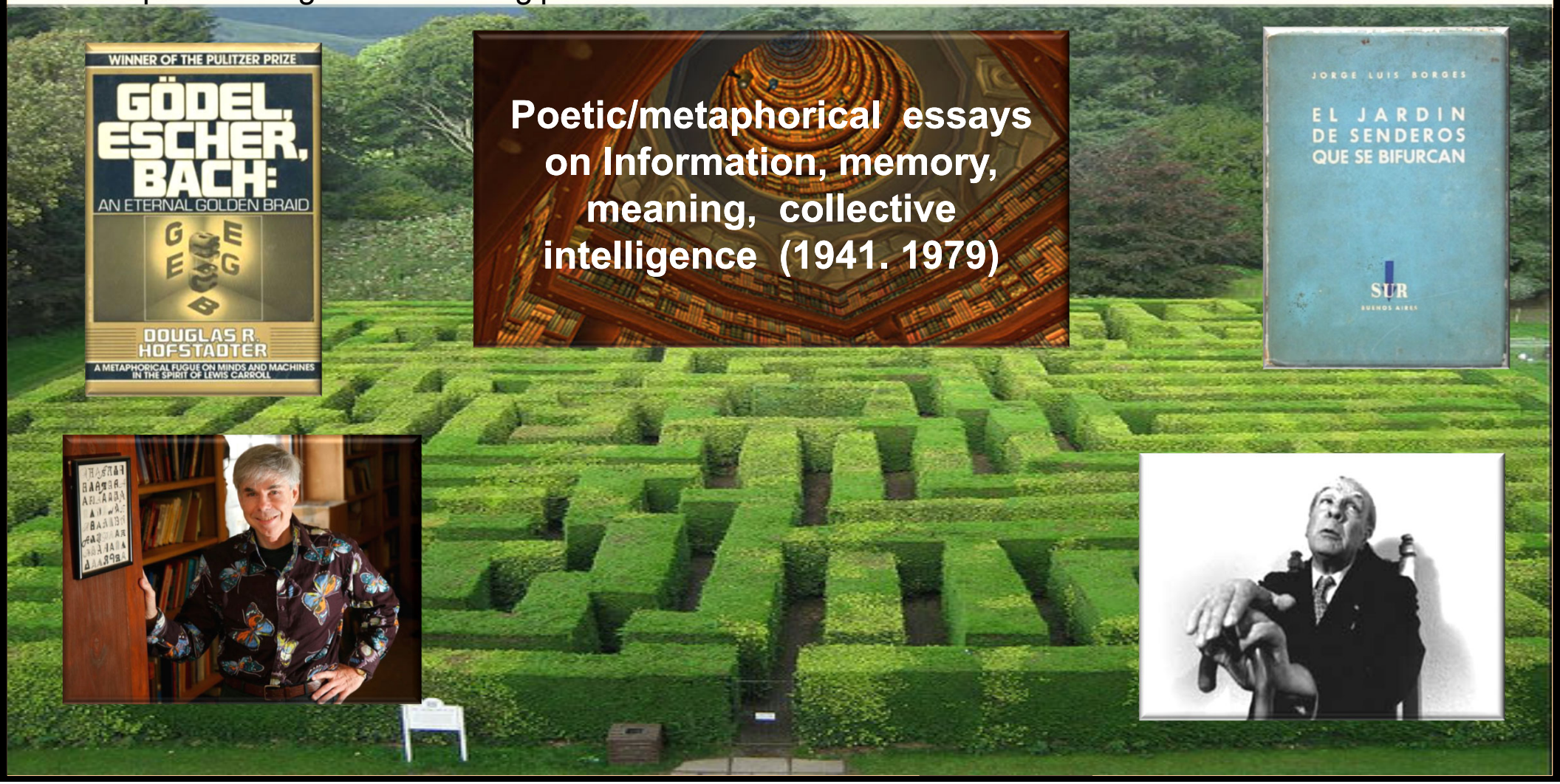
more upcoming readings (check brightspace)

- Paper Presentation: 20%
  - Present (501) and lead (501&440) the discussio
    - Enginet students post/send video or join by Zoon
- Module 2: Systems Science
  - Presenter 5 :
    - Klir, G.J. [2001]. *Facets of systems Scienc*
      - Optional:
      - Rosen, R. [1986]. "Some comments on systen *Science*. Springer. pp: 241-243.
      - Wigner, E.P. [1960], "The unreasonable effecti York University, May 11, 1959. *Comm. Pure*
  - Presenter 6:
    - Klir, G.J. [2001]. *Facets of systems Scienc*
  - Presenter 7:
    - Klir, G.J. [2001]. *Facets of systems Scienc*
      - Optional: Klir, G.J. [2001]. *Facets of system*
  - Presenter 8:
    - Schuster, P. (2016). The end of Moore's law: **21**(S1): 6-9. DOI 10.1002/cplx.21824.
  - Presenter 9:
    - Von Foerster, H., P. M. Mora and L. W. Amic
- Future Modules
  - See brightspace

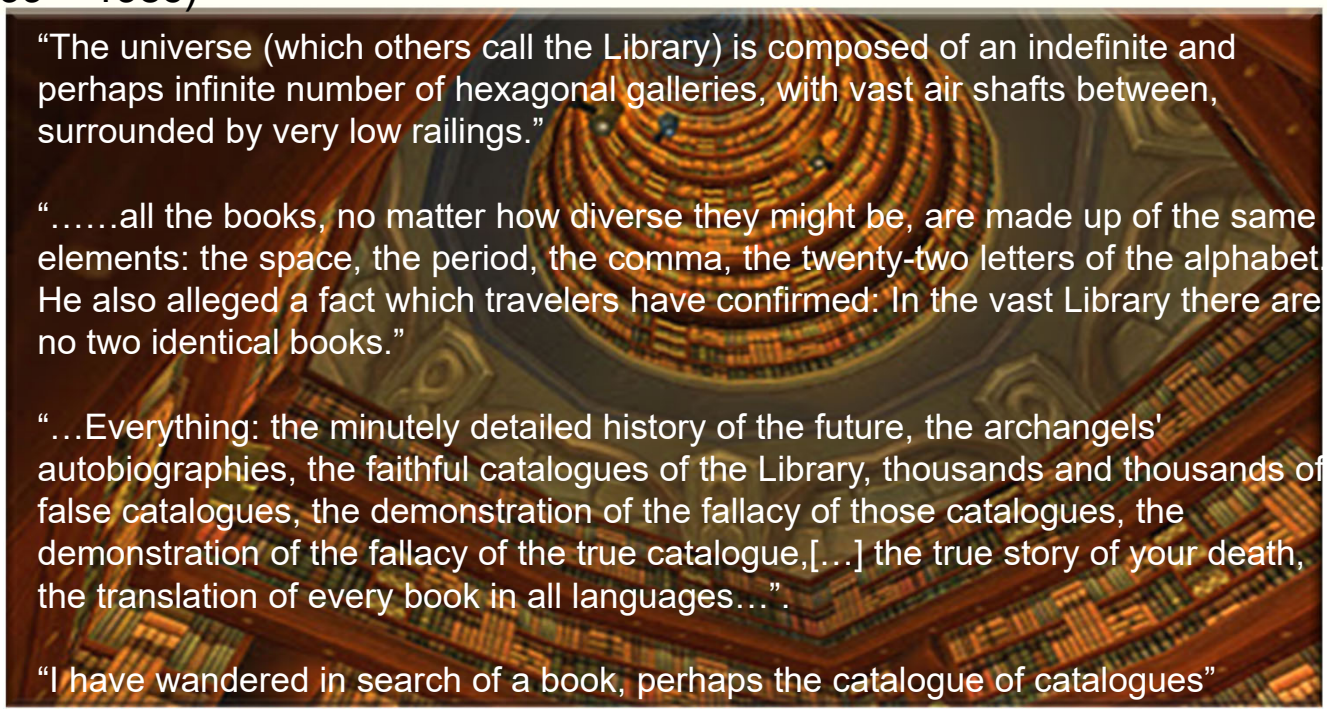


BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK
rocha@indiana.edu
informatics.indiana.edu/rocha

Personal path in the garden of forking paths



Jorge Luis Borges (1899 – 1986)

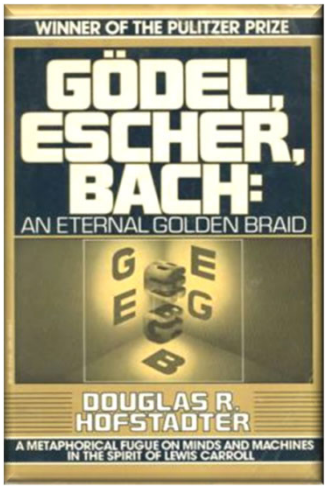


"The universe (which others call the Library) is composed of an indefinite and perhaps infinite number of hexagonal galleries, with vast air shafts between, surrounded by very low railings."

"……all the books, no matter how diverse they might be, are made up of the same elements: the space, the period, the comma, the twenty-two letters of the alphabet. He also alleged a fact which travelers have confirmed: In the vast Library there are no two identical books."

"…Everything: the minutely detailed history of the future, the archangels' autobiographies, the faithful catalogues of the Library, thousands and thousands of false catalogues, the demonstration of the fallacy of those catalogues, the demonstration of the fallacy of the true catalogue,[…] the true story of your death, the translation of every book in all languages…".

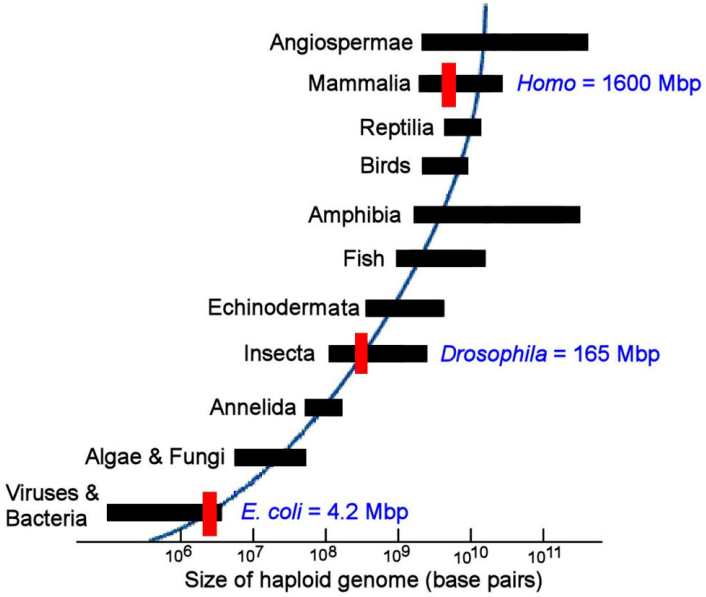
"I have wandered in search of a book, perhaps the catalogue of catalogues"



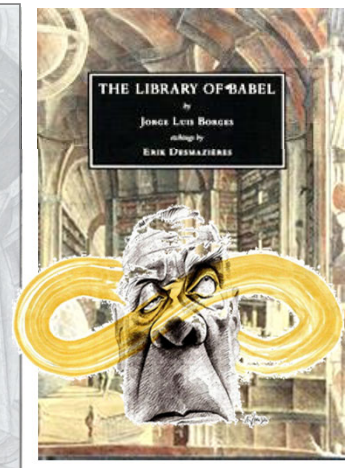
Poetic essays on information and memory (1941)

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
informatics.indiana.edu/rocha/academics/i-bic

numbers



- Each book
  - 25 characters, written in any sequence for 410 pages of 40 lines of 80 characters
  - 410*40*80 = sequence of ≈ $10^{6.1}$ characters
    - ≈ $10^{7.2}$ base pairs (10 Mbp)
    - ≈ 1 book to store E.Coli genotype, 10 for drosophila, and 100 for human
- How many possible books?
  - = $25^{(410*40*80)}$ combinations = $25^{1,312,000}$ books!
    - ≈ 1.956 x $\mathbf{10^{1,834,097}}$ books
  - Total number of atoms in the current, observable universe is about 10^{80}
    - If each book were the size of an atom, library would hold $10^{1,834,017}$ **universes**!
  - Yet finite!
    - Can also be reproduced with just two symbols (cf Quine, Turing, Leibniz)

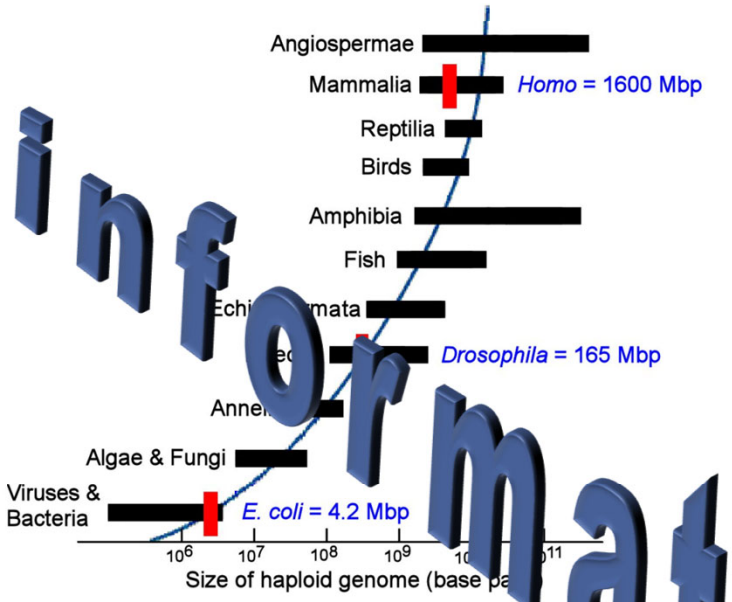


Information Space Is finite but larger than Physical space

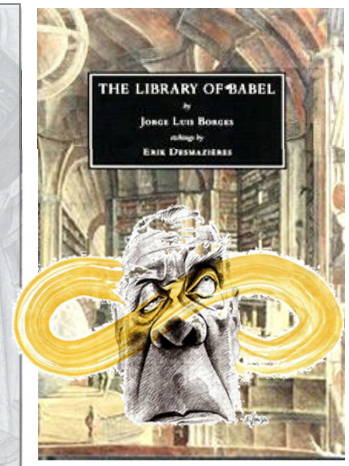
"the Library is so enormous that any reduction of human origin is infinitesimal."
"every copy is unique, irreplaceable, but (since the Library is total) there are always several hundred thousand imperfect facsimiles: works which differ only in a letter or a comma."

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
informatics.indiana.edu/rocha/academics/i-bic

numbers



- Each book
  - 25 characters, written in any sequence for 410 pages of 40 lines of 80 characters
  - 410*40*80 = sequence of ≈ $10^{6.1}$ characters
    - ≈ $10^{7.2}$ base pairs (10 Mbp)
    - ≈ 1 book to store E.Coli genotype, 10 for drosophila, and 100 for human
- How many possible books?
  - = $25^{(410*40*80)}$ combinations = $25^{1,312,000}$ books!
    - ≈ 1.956 x $\mathbf{10^{1,834,097}}$ books
  - Total number of atoms in the current, observable universe is about 10^{80}
    - If each book were the size of an atom, library would hold $10^{1,834,017}$ **universes**!
  - Yet finite!
    - Can also be reproduced with just two symbols (cf Quine, Turing, Leibniz)

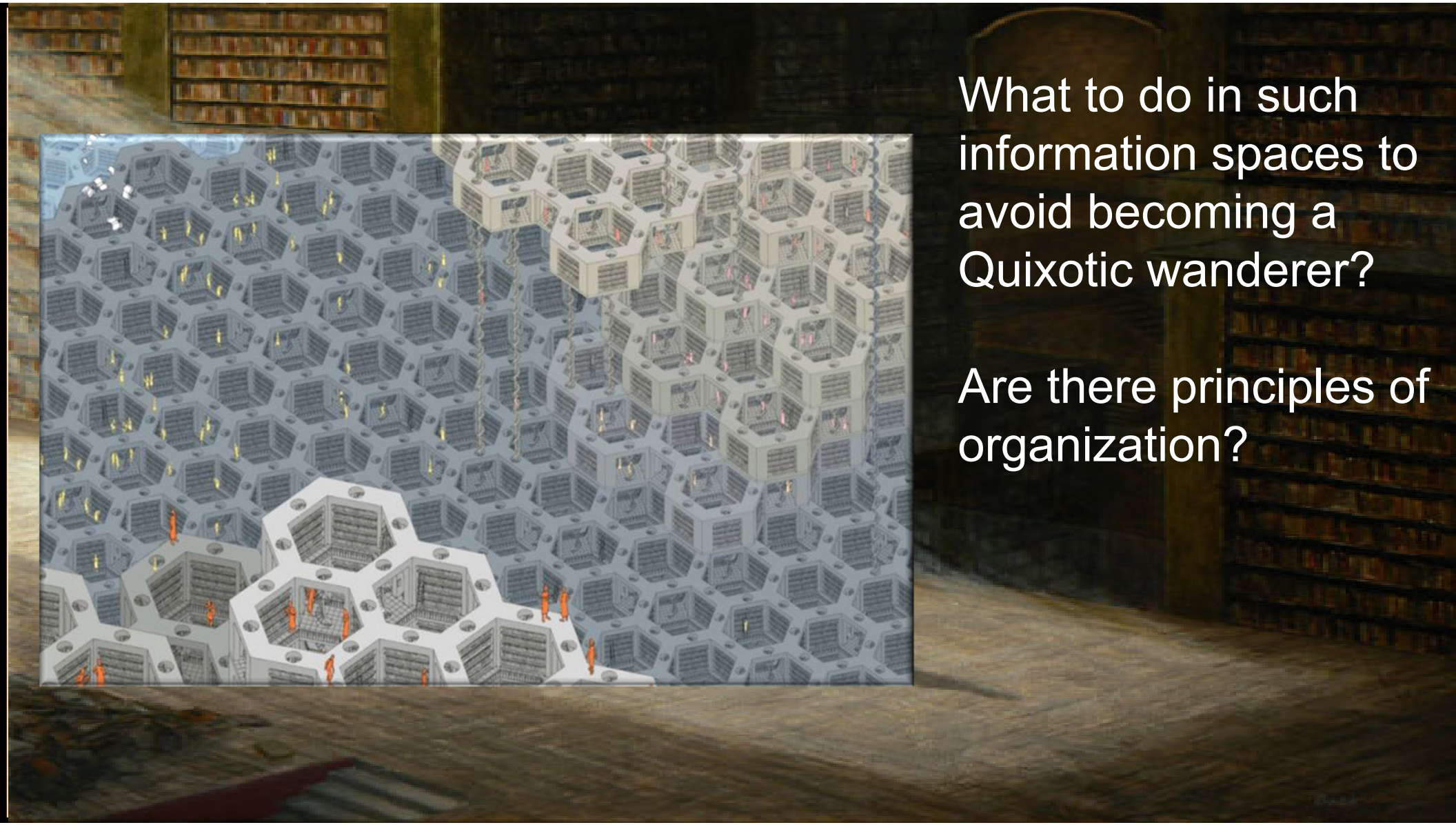


Information Space Is finite but larger than Physical space

"the Library is so enormous that any reduction o[f] human origin is infinitesimal."
"every copy is unique, irreplaceable, but (since the library is total) there are always several hundred thousand imperfect facsimiles: works which differ only in a letter or a comma."

information

BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK
rocha@binghamton.edu
informatics.indiana.edu/rocha/academics/i-bic


What to do in such information spaces to avoid becoming a Quixotic wanderer?
Are there principles of organization?

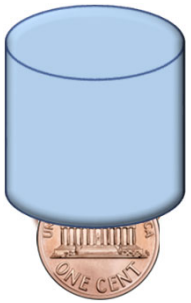

How did we get here?

observer and choice

- *Information* is defined as "a measure of the freedom from *choice* with which a message is *selected* from the set of all possible messages"
- Bit (short for *binary digit*) is the most elementary **choice** one can make
  - Between two items: "0' and "1", "heads" or "tails", "true" or "false", etc.
  - Bit is equivalent to the choice between two equally likely alternatives
    - Example, if we know that a coin is to be tossed, but are unable to see it as it falls, a message telling whether the coin came up heads or tails gives us one bit of information

1 Bit of *information*

uncertainty ***removed***, information gained



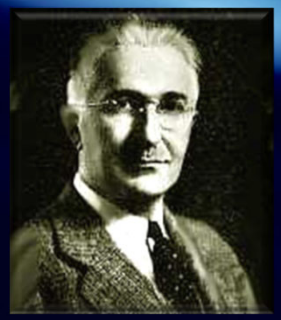
1 Bit of uncertainty

H,T?

choice between 2 symbols recognized by an observer

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

rocha@indiana.edu
informatics.indiana.edu/rocha



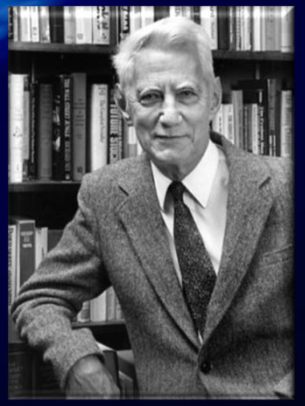
- Information is transmitted through noisy communication channels
  - Ralph Hartley and Claude Shannon (at Bell Labs), the fathers of Information Theory, worked on the problem of efficiently transmitting information; i. e. ***decreasing the uncertainty*** in the transmission of information.

Hartley, R.V.L., "Transmission of Information", *Bell System Technical Journal*, July 1928, p.535.

C. E. Shannon [1948], "A mathematical theory of communication". *Bell System Technical Journal*, **27**:379-423 and 623-656

C. E. Shannon, "A Symbolic analysis of relay and switching circuits" .*MS Thesis*, (unpublished) MIT, 1937.

C. E. Shannon, "An algebra for theoretical genetics." *Phd Dissertation*, MIT, 1940.



- Multiplication Principle
  - "If some choice can be made in M different ways, and some subsequent choice can be made in N different ways, then there are M x N different ways these choices can be made in succession" [Paulos]
    - 3 shirts and 4 pants = 3 x 4 = 12 outfit choices

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK
rocha@indiana.edu
informatics.indiana.edu/rocha

- Nonspecificity
  - Hartley measure
    - The amount of uncertainty associated with a set of alternatives (e.g. messages) is measured by the **amount of information needed to remove the uncertainty**

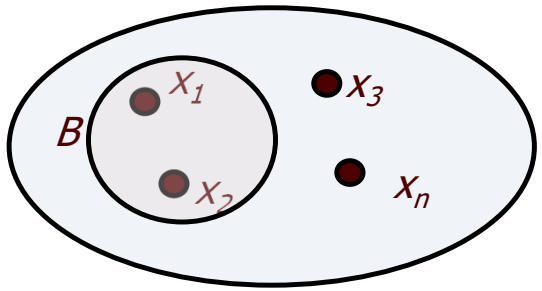
Quantifies how many yes-no questions need to be asked to establish what the correct alternative is

Elementary Choice is between 2 alternatives: 1 bit

$$H(B) = \log_2(2) = 1$$

$$\log_2(4) = 2 \qquad 2^2 = 4$$

A = Set of Alternatives



$$H(A) = \log_2 |A|$$

Measured in bits

Number of Choices

$$\log_2(16) = 4 \qquad 2^4 = 16$$

$$\log_2(1) = 0$$

$$H(A) = \log_2(16) = 4$$

$$H(B) = \log_2(4) = 2$$

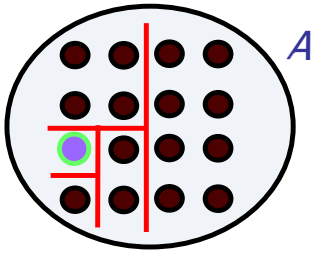
$$H(A) = \log_2|A|$$

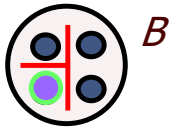
Measured in bits

Number of Choices

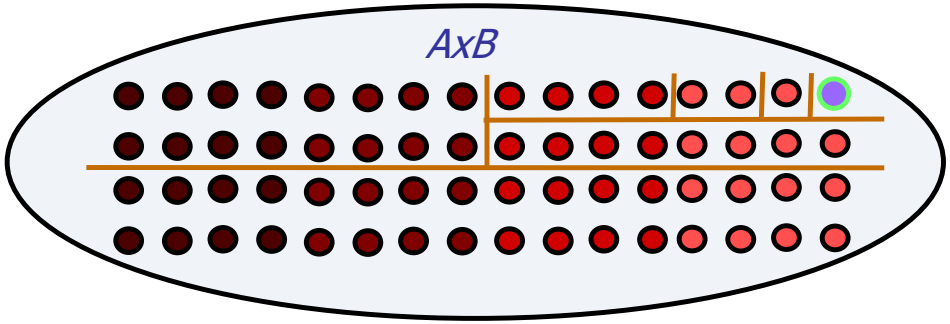
Quantifies how many yes-no questions need to be asked to establish what the correct alternative is

- Example
  - Menu Choices
    - A = 16 Entrees
    - B = 4 Desserts
  - How many dinner combinations?
    - 16 x 4 = 64

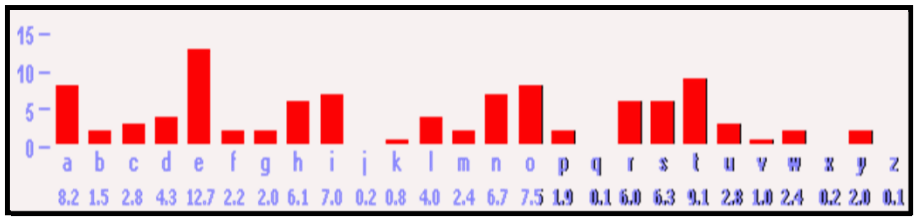


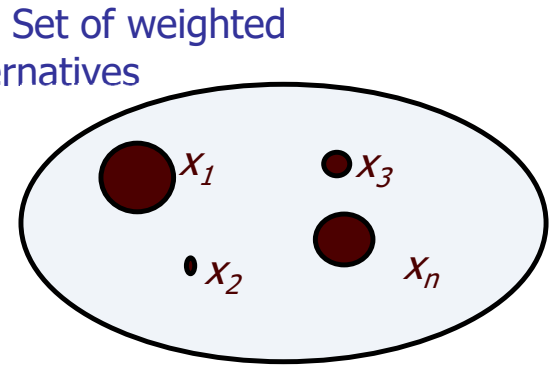


$$H(A \times B) = \log_2(16 \times 4) =$$
$$= \log_2(16) + \log_2(4) = 6$$



uncertainty-based information





- Shannon's measure
  - The ***average*** amount of uncertainty associated with a set of ***weighted*** alternatives (e.g. messages) is measured by the ***average*** amount of information needed to remove the uncertainty

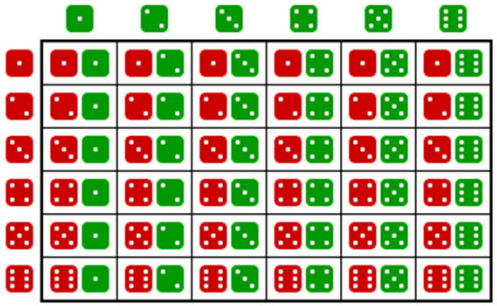
$$H_S(A) = -\sum_{i=1}^{n} p(x_i)\log_2\big(p(x_i)\big)$$

Measured in bits

Probability of alternative

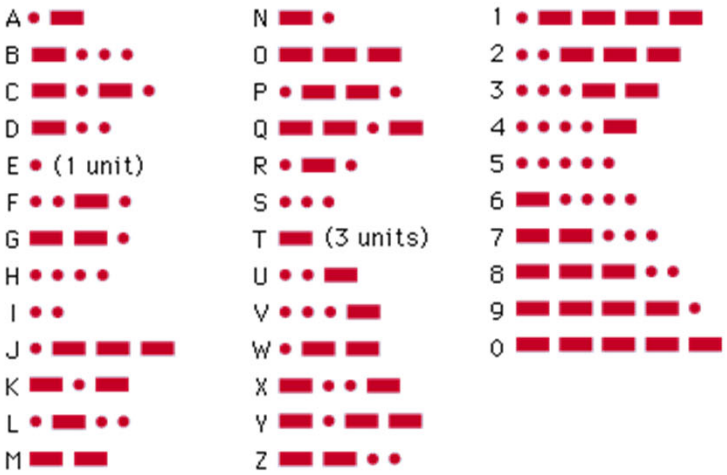
BINGHAMTON UNIVERSITY STATE UNIVERSITY OF NEW YORK
rocha@indiana.edu
informatics.indiana.edu/rocha

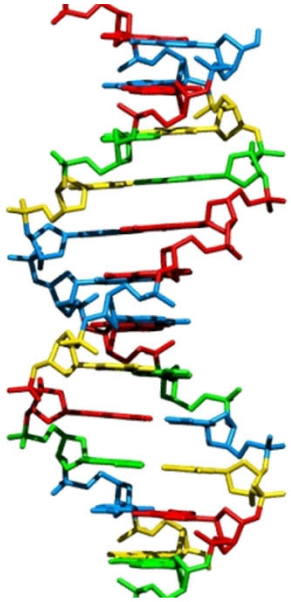
alphabet examples

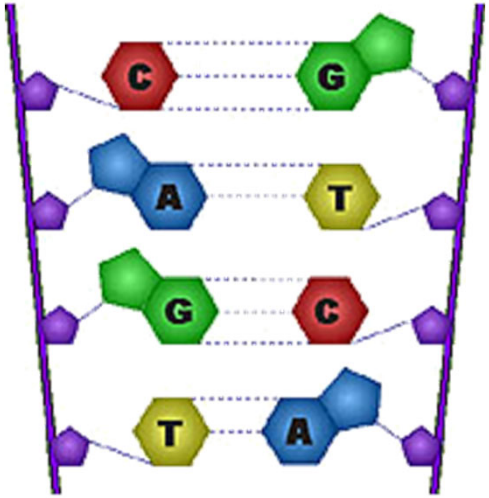


Message encoded in an alphabet of ***n*** symbols, for example:

- English (26 letters + space + punctuations)
- Morse code (dot, dash, space)
- DNA (A, T, G, C)
- Two dice (11 integers)





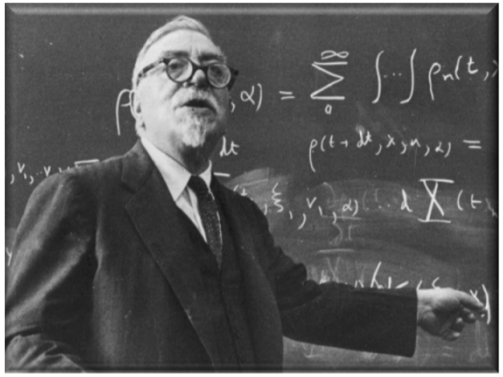


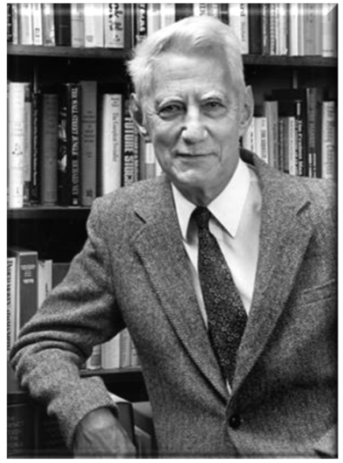
BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

rocha@indiana.edu
informatics.indiana.edu/rocha

readings

- Class Book
  - Klir, G.J. [2001]. *Facets of systems science*. Springer.
- Papers and other materials
  - Presenter 1
    - Heims, S.G. [1991]. *The Cybernetics Group*. MIT Press. Chapters: 1 - 2.
      - Optional: Chapters 11-12.
      - Optional: McCulloch, W. and W. Pitts [1943], "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics* **5**:115-133.
  - Presenter 2
    - Gleick, J. [2011]. *The Information: A History, a Theory, a Flood*. Random House. Chapter 8.
      - Optional: Prokopenko, Mikhail, Fabio Boschetti, and Alex J. Ryan. "An information theoretic primer on complexity, self-organization, and emergence." *Complexity* **15**.1 (2009): 11-28.
  - Presenter 3
    - Brenner, Sydney. [2012]. "History of Science. The Revolution in the Life Sciences". *Science* **338** (6113): 1427-8.
    - Brenner, Sydney. [2012]. "Turing centenary: Life's code script. *Nature* **482** (7386) (February 22): 461-461.
    - Cobb, Matthew. [2013]. "1953: When Genes Became 'Information'." *Cell* **153** (3): 503-506.
      - Optional: Searls, David B. [2010]. "The Roots of Bioinformatics". *PLoS Computational Biology* **6**(6): e1000809.
  - Presenter 4
    - Weaver, W. [1948]. "Science and Complexity". *American Scientist*, **36**(4): 536-44. Also available in Klir, G.J. [2001]. *Facets of systems Science*. Springer, pp: 533-540.





BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

rocha@indiana.edu
informatics.indiana.edu/rocha