



Block 1: Intro to the Course

ELEC 573: Network Science and Analytics

Santiago Segarra

Electrical and Computer Engineering

Rice University

segarra@rice.edu

Fall 2021



Course logistics

Networks: What are they and why are they important?

Data Science for Networks: A bit of what we do

To link or not to link: Shakespeare and the networks

Who am I, where to find me, lecture times



- ▶ Santiago Segarra
- ▶ Assistant Professor, Dept. of Electrical and Computer Engineering
- ▶ Duncan Hall 2047, segarra@rice.edu
- ▶ <http://segarra.rice.edu/>
- ▶ Where? We meet in Howard Keck Hall 107 starting September 13th
 - ⇒ In August 30th we meet remotely and September 6th is labor day
- ▶ When? Mondays 4:00 pm to 6:50 pm
- ▶ My weekly office hours, Mondays at 3:00 pm and after class
 - ⇒ Starting August 30th
 - ⇒ Both in zoom and in person
- ▶ All important materials and announcements will be posted in Canvas



- ▶ Boning Li
- ▶ 4th year Ph.D. student, Dept. of Electrical and Computer Engineering
- ▶ boning.li@rice.edu

- ▶ Madeline Navarro
- ▶ 2nd year Ph.D. student, Dept. of Electrical and Computer Engineering
- ▶ mn51@rice.edu

- ▶ They took the course in the past ⇒ Liked it enough to be graders!
- ▶ Each will set 1 hour slots of office hours ⇒ Inform via Canvas



(I) Probability theory and linear algebra

- ▶ Random variables, distributions, expectations, Markov processes
- ▶ Vector/matrix notation, systems of linear equations, eigenvalues



(I) Probability theory and linear algebra

- ▶ Random variables, distributions, expectations, Markov processes
- ▶ Vector/matrix notation, systems of linear equations, eigenvalues

(II) Statistical inference

- ▶ Statistical inference useful to “learn” from network data
- ▶ Estimators, MSE, bias vs. variance



(I) Probability theory and linear algebra

- ▶ Random variables, distributions, expectations, Markov processes
- ▶ Vector/matrix notation, systems of linear equations, eigenvalues

(II) Statistical inference

- ▶ Statistical inference useful to “learn” from network data
- ▶ Estimators, MSE, bias vs. variance

(III) Programming

- ▶ Will use Python for homework and your project
- ▶ You can use any network analysis package you prefer
- ▶ Check the NetworkX package for Python

Homework, project and grading



- (I) 3 homework sets worth 40% (plus an ungraded Homework 0)
 - ▶ Mix of analytical problems and programming assignments
 - ▶ Collaboration accepted, welcomed, and encouraged
 - ▶ However, the submitted work **must** be your own



Homework, project and grading

(I) 3 homework sets worth 40% (plus an ungraded Homework 0)

- ▶ Mix of analytical problems and programming assignments
- ▶ Collaboration accepted, welcomed, and encouraged
- ▶ However, the submitted work **must** be your own

(II) Research project on a topic of your choice, worth 60%

- ▶ Important and demanding part of this class. Three deliverables:
 - 1) **Proposal** by the end of week 9, worth 10%
 - 2) **Progress report** by the end of week 12, worth 15%
 - 3) **Final report and recorded presentation**, worth 35%

Homework, project and grading



(I) 3 homework sets worth 40% (plus an ungraded Homework 0)

- ▶ Mix of analytical problems and programming assignments
- ▶ Collaboration accepted, welcomed, and encouraged
- ▶ However, the submitted work **must** be your own

(II) Research project on a topic of your choice, worth 60%

- ▶ Important and demanding part of this class. Three deliverables:

- 1) **Proposal** by the end of week 9, worth 10%
- 2) **Progress report** by the end of week 12, worth 15%
- 3) **Final report and recorded presentation**, worth 35%

- ▶ This is a research-oriented graduate level class

⇒ Focus should be on thinking, reading, asking, implementing



Overall syllabus (tentative) and deadlines (set)

Wk.	Date	Topic	HW	Project
1	23-Aug	Introduction to course	HW0 out	
2	30-Aug	Graph theory / Centrality measures	HW0 solutions posted	
3	6-Sep	LABOR DAY (no class)	HW1 out	
4	13-Sep	Centrality measures / Community detection		
5	20-Sep	Community detection		
6	27-Sep	Signal Processing and Deep learning for graphs	HW1 due	
7	4-Oct	Signal Processing and Deep learning for graphs	HW2 out	
8	11-Oct	FALL BREAK (no class)		
9	18-Oct	Network models	HW2 due	
10	25-Oct	Network models	HW3 out	Project proposal due
11	1-Nov	Epidemics		
12	8-Nov	Inference of network topologies, features, and processes	HW3 due	
13	15-Nov	Inference of network topologies, features, and processes		
14	22-Nov	Inference of network topologies, features, and processes		Project progress report
15	29-Nov	Inference of network topologies, features, and processes		
	13-Dec	Project presentation (video recording) and final report due		

A few things about handing-in homework

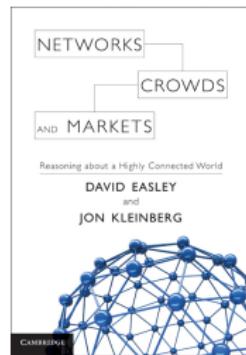
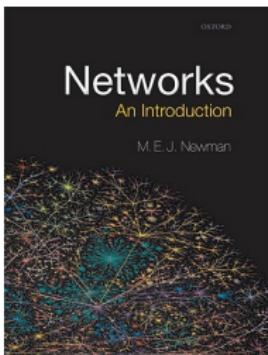
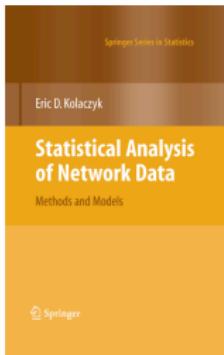


- ▶ All submissions **must** be via Canvas
- ▶ Can be scanned copies of handwritten work if you are tidy
- ▶ All homework is released on **Monday evenings**
- ▶ All homework is due on **Tuesday by midnight**
- ▶ For coding exercises, ready-to-run code must be included
 - ⇒ Submit all your work in a compressed folder
 - ⇒ Jupyter notebooks highly appreciated
 - ⇒ 'lastname_homework_i'
- ▶ If anything comes up, please come talk to me **with time!**
 - ⇒ An ounce of prevention is worth a pound of cure



Reading material

- ▶ We will use **lecture slides** to cover the material
 - ⇒ Research papers, tutorials also posted on canvas
- ▶ Some books for your reference:
- ▶ Eric D. Kolaczyk, "*Statistical Analysis of Network Data: Methods and Models,*" Springer
- ▶ M. E. J. Newman, "*Networks: An Introduction,*" Oxford U. Press
- ▶ D. Easley and J. Kleinberg, "*Networks, Crowds, and Markets: Reasoning About a Highly Connected World,*" Cambridge U. Press





- ▶ I **work hard** for this course, expect you to do the same
 - ✓ Come to class, be on time, pay attention, ask
 - ✓ Check out the additional suggested readings
 - ✓ Play with network analysis software
 - ✓ Search for datasets
 - ✓ Do all of your homework
 - ✗ Do not hand in as yours the solution of others



- ▶ I **work hard** for this course, expect you to do the same
 - ✓ Come to class, be on time, pay attention, ask
 - ✓ Check out the additional suggested readings
 - ✓ Play with network analysis software
 - ✓ Search for datasets
 - ✓ Do all of your homework
 - ✗ Do not hand in as yours the solution of others
- ▶ **Come and learn!**

Sources



- ▶ ECE442 from U. Rochester by Gonzalo Mateos (collaborator)
- ▶ 1.022 from MIT by Ali Jadbabie (co-developer)
- ▶ 6.207 from MIT by Devavrat Shah
- ▶ 18.06 from MIT by Gilbert Strang
- ▶ AMTH 462 from Yale by Dan Spielman
- ▶ PHYS 5116 from NorthEastern by Albert-László Barabási
- ▶ ESE 303 from U. Pennsylvania by Alejandro Ribeiro (former advisor)

Taking, auditing, and coming to the course



- ▶ You should fall under one of these three categories
- ▶ You are **taking** the course (majority)
 - ⇒ Homework and final project
- ▶ You are **auditing** the course (AUD grade or similar)
 - ⇒ At least 2 out of 3 homeworks (not counting homework 0)
- ▶ You are just **coming** to the course
 - ⇒ Come and listen, no strings attached, ask for material
- ▶ If you do not fall in any of these categories, let me know

Networks: What and why?



Course logistics

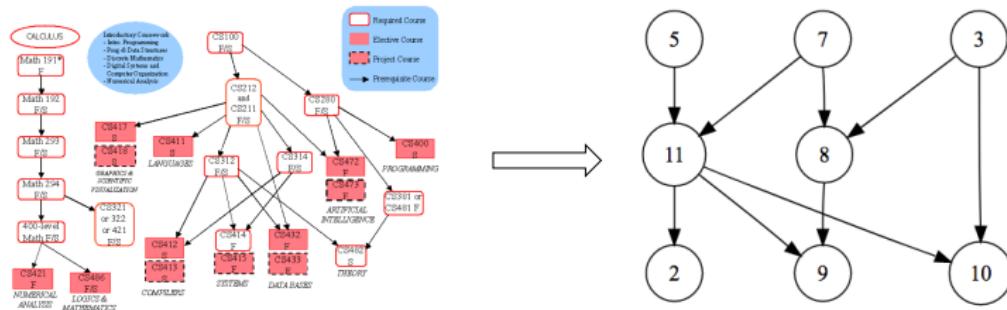
Networks: What are they and why are they important?

Data Science for Networks: A bit of what we do

To link or not to link: Shakespeare and the networks



- ▶ Pattern of interconnections among a set of things
- ▶ Very general definition ⇒ enormous range of topics
- ▶ Growing public fascination with connectedness of modern society

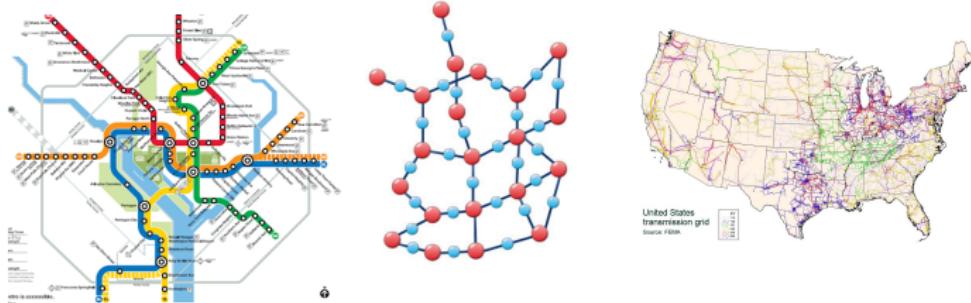


- ▶ Complex system modeled as a mathematical network (graph)
- ▶ Understanding systems at this level of abstraction
⇒ Powerful tool across disciplines



Increasing interest

- ▶ Network-based analysis has a long-standing history
 - ⇒ Laws of electrical circuitry (Kirchoff, 1845)
 - ⇒ Molecular structure in chemistry (Cayley, 1874)
 - ⇒ Power grids (1910), telecommunications and the Internet (1960)

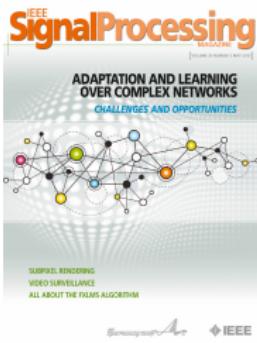


- ▶ **Explosion of interest** in the last two decades
 - ⇒ Systems-level perspective in science
 - ⇒ High-throughput data collection and computational power
 - ⇒ Globalization and connected modern society



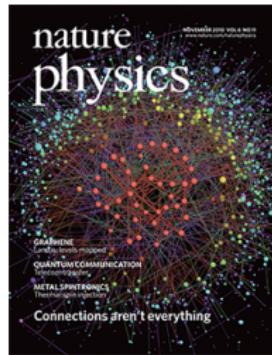
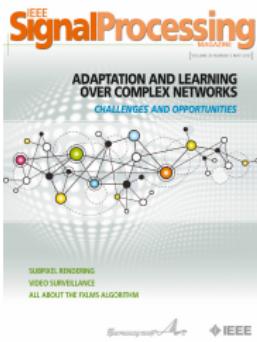
Network Science

- ▶ Study of complex systems through their network representations
Ex: economy, metabolism, brain, society, Web, ...
- ▶ Universal language for describing complex systems and data
 - ▶ Striking similarities in networks across science, nature, technology
- ▶ Shared vocabulary across fields, cross-fertilization
 - ▶ From biology to physics, economics to statistics, CS to sociology





- ▶ Study of complex systems through their network representations
Ex: economy, metabolism, brain, society, Web, ...
- ▶ Universal language for describing complex systems and data
 - ▶ Striking similarities in networks across science, nature, technology
- ▶ Shared vocabulary across fields, cross-fertilization
 - ▶ From biology to physics, economics to statistics, CS to sociology



- ▶ **Impact:** social networking, drug design, smart infrastructure, ...



Economic impact

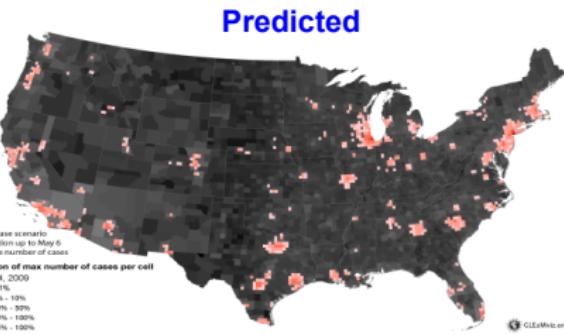
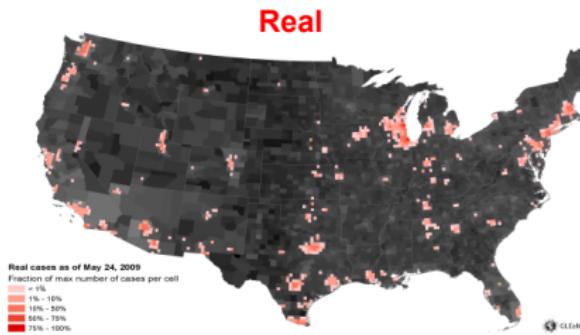
- ▶ Google
Market cap:
\$1070 billion
- ▶ Facebook
Market cap:
\$766 billion
- ▶ Cisco
Market cap:
\$179 billion
- ▶ Apple
Market cap:
\$2020 billion



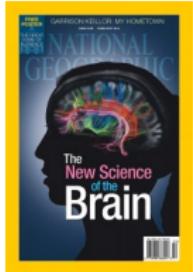
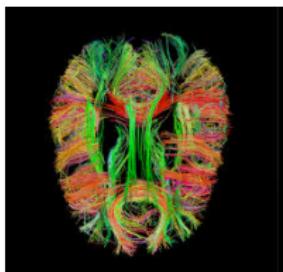


Healthcare impact

- ▶ Prediction of **epidemics**, e.g. the 2009 H1N1 pandemic



- ▶ Human Connectome Project to map-out **brain** circuitry





- ▶ What are the **goals** of Network Science?
 - ▶ **Reveal** patterns and statistical properties of network data
 - ▶ **Understand** the underpinnings of network behavior and structure
 - ▶ **Engineer** more resource-efficient, robust, socially-intelligent networks



- ▶ What are the **goals** of Network Science?
 - ▶ **Reveal** patterns and statistical properties of network data
 - ▶ **Understand** the underpinnings of network behavior and structure
 - ▶ **Engineer** more resource-efficient, robust, socially-intelligent networks
- ▶ **Characteristics:** interdisciplinary, empirical, quantitative, computational



- ▶ What are the **goals** of Network Science?
 - ▶ Reveal patterns and statistical properties of network data
 - ▶ Understand the underpinnings of network behavior and structure
 - ▶ Engineer more resource-efficient, robust, socially-intelligent networks
- ▶ **Characteristics:** interdisciplinary, empirical, quantitative, computational
- ▶ **Empirical** study of graph-valued data to find patterns and principles
 - ▶ Collection, measurement, summarization, visualization?



- ▶ What are the **goals** of Network Science?
 - ▶ Reveal patterns and statistical properties of network data
 - ▶ Understand the underpinnings of network behavior and structure
 - ▶ Engineer more resource-efficient, robust, socially-intelligent networks
- ▶ **Characteristics:** interdisciplinary, empirical, quantitative, computational
- ▶ **Empirical** study of graph-valued data to find patterns and principles
 - ▶ Collection, measurement, summarization, visualization?
- ▶ Mathematical **models**. Graph theory meets statistical inference
 - ▶ Understand, predict, discern nominal vs anomalous behavior?



- ▶ What are the **goals** of Network Science?
 - ▶ Reveal patterns and statistical properties of network data
 - ▶ Understand the underpinnings of network behavior and structure
 - ▶ Engineer more resource-efficient, robust, socially-intelligent networks
- ▶ **Characteristics:** interdisciplinary, empirical, quantitative, computational
- ▶ **Empirical** study of graph-valued data to find patterns and principles
 - ▶ Collection, measurement, summarization, visualization?
- ▶ Mathematical **models**. Graph theory meets statistical inference
 - ▶ Understand, predict, discern nominal vs anomalous behavior?
- ▶ **Algorithms** for graph analytics
 - ▶ Computational challenges, scalability, tractability vs optimality?



Examples of networks

- ▶ Network analysis spans the sciences, humanities and arts
- ▶ Let's see a few examples from four general areas
 - ▶ Technological
 - ▶ Biological
 - ▶ Social
 - ▶ Information
- ▶ Standard taxonomy, by no means the only one
 - ⇒ “Soft” classification, networks may fall in multiple categories



Technological networks

- Ex: communication, transportation, energy, sensor networks

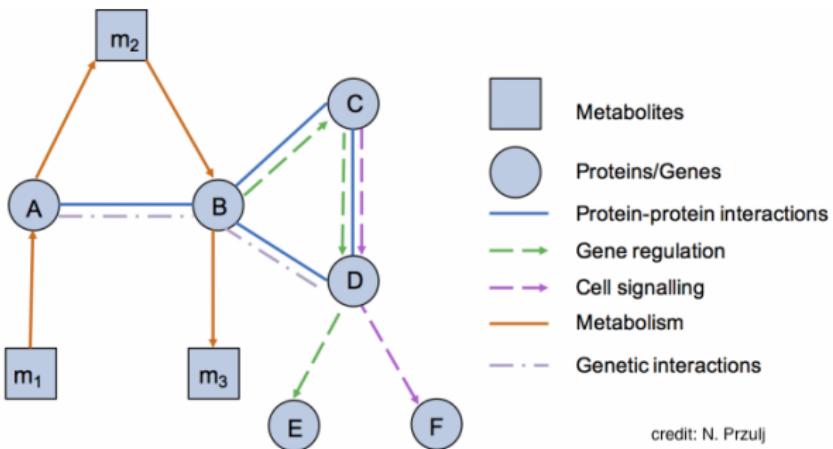


- Q1: What does the Internet look like today? How big is it?
- Q2: How will the traffic from New York to Chicago look tomorrow?
- Q3: How can we unveil anomalous traffic patterns?

Biological networks



- Ex: neurons, gene regulatory, protein interaction, metabolic paths, predator-prey, ecological networks



- Q1: Are certain gene interactions more common than expected?
- Q2: Which parts of the brain “communicate” during a given task?
- Q3: Can we predict biological function of proteins from interactions?



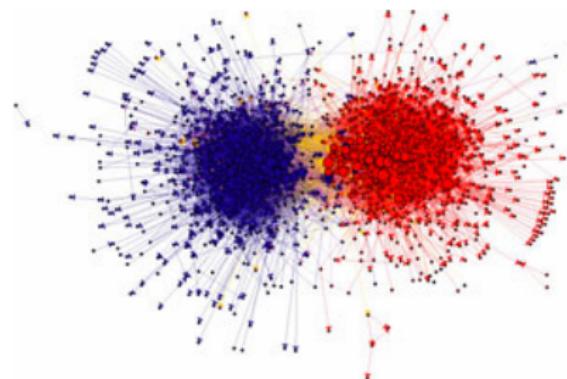
- ▶ **Ex:** friendship, corporate, email exchange, international relations, financial networks



- ▶ **Q1:** What are the mechanisms underpinning friendship formation?
- ▶ **Q2:** Which actors are central to the network and which peripheral?
- ▶ **Q3:** Can we identify overlapping communities?



- Ex: WWW, Twitter, co-citation between academic journals, blogosphere, paper co-authorship, peer-to-peer networks



- Q1: How does the size and structure of the WWW change in time?
- Q2: How can we use network analysis for authorship attribution?
- Q3: Can we track information cascades in online social media?

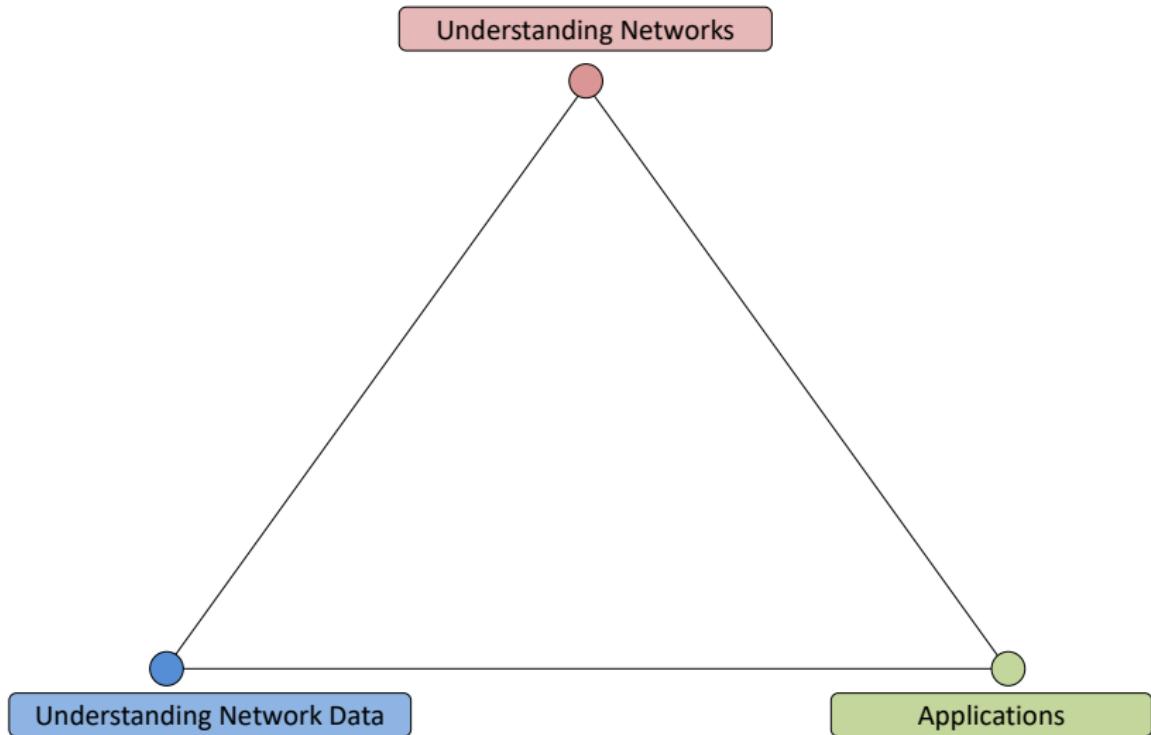


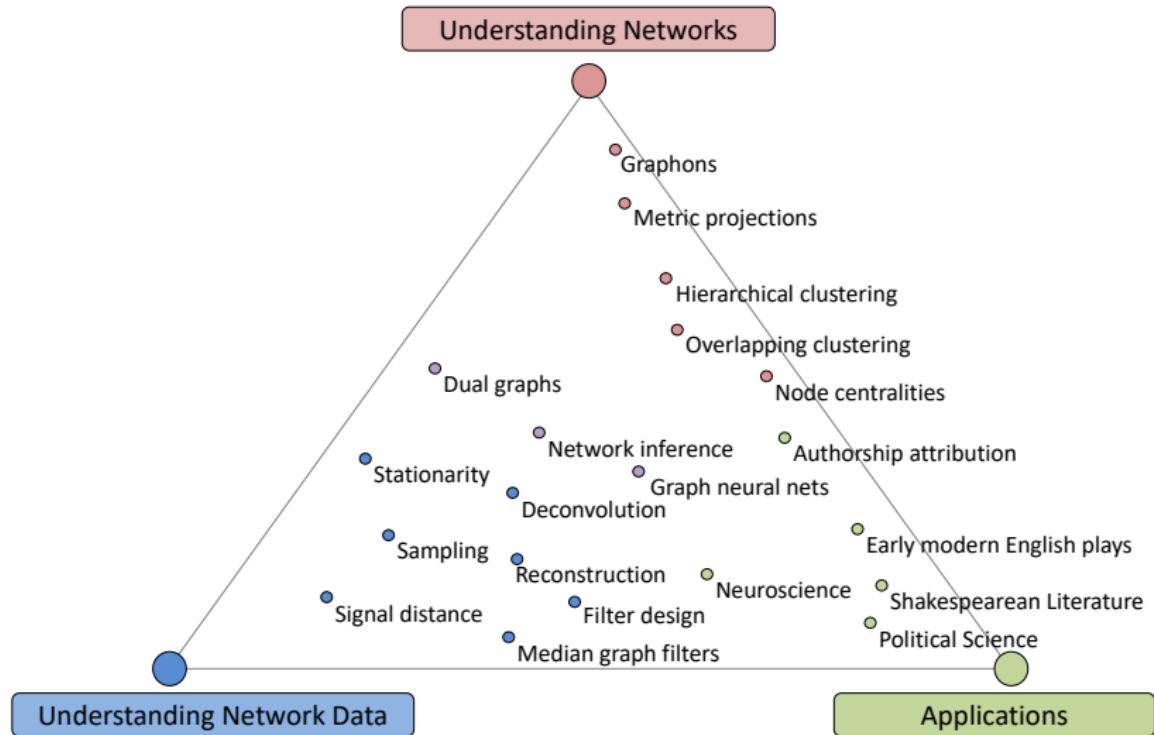
Course logistics

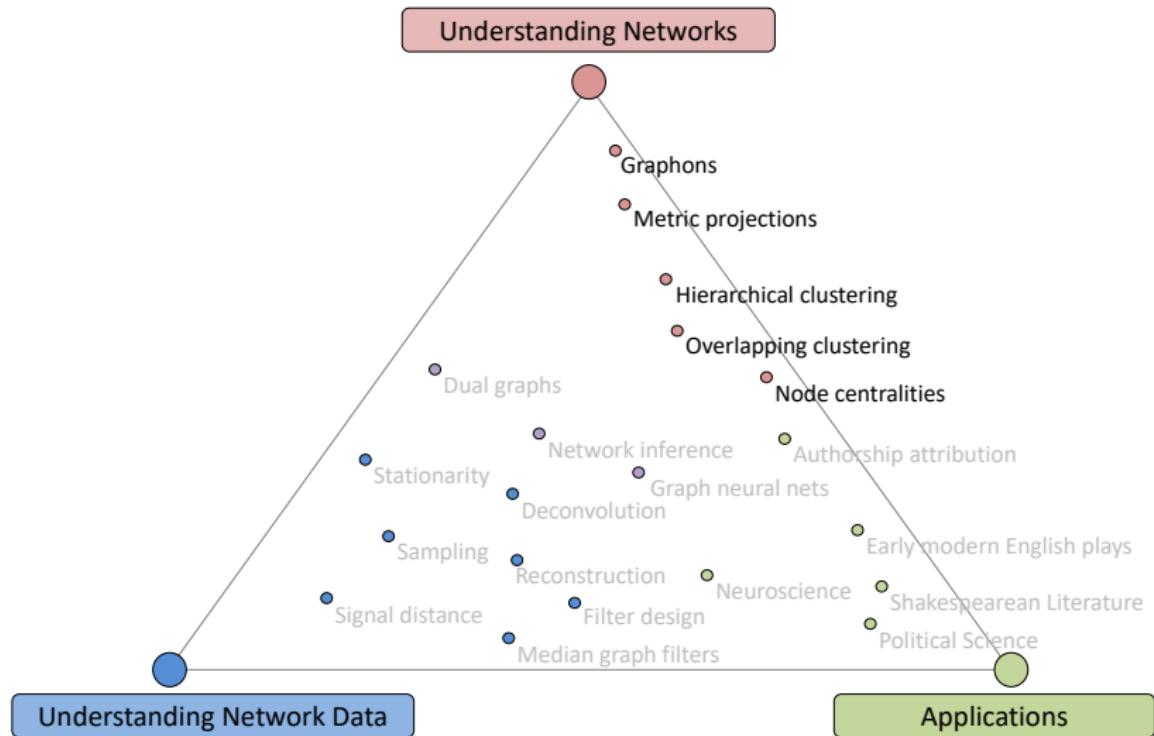
Networks: What are they and why are they important?

Data Science for Networks: A bit of what we do

To link or not to link: Shakespeare and the networks



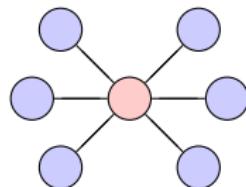






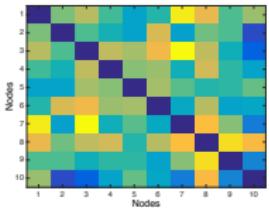
Node centralities

- ▶ Identify the **most important nodes** in a graph given its **topology**
 - ⇒ Not based on the **nature** of the particular node
- ▶ Different definitions of **importance** give rise to many **centrality measures**
 - ⇒ Degree, closeness, eigenvector, betweenness
 - ⇒ They induce a **centrality ranking** on the nodes
- ▶ Centrality measures are **widely used**
 - ⇒ Targeted marketing
 - ⇒ Network vulnerability to attacks
 - ⇒ Epidemiology control
 - ⇒ Power in exchange networks





Clustering



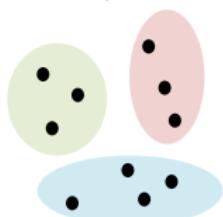
Weighted dissimilarity network

- Large dimension
- No structure on weights



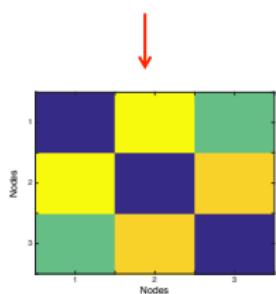
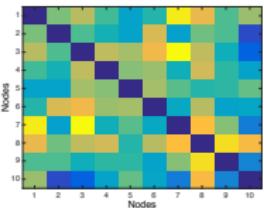
Clustering

Reduce dimensionality



Weighted dissimilarity network

- Large dimension
- No structure on weights

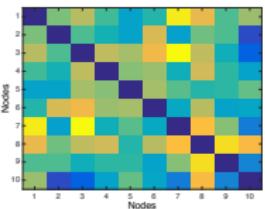
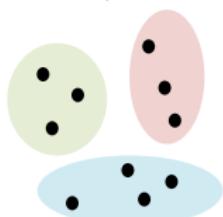


Smaller network

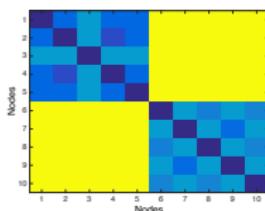
Clustering



Reduce dimensionality

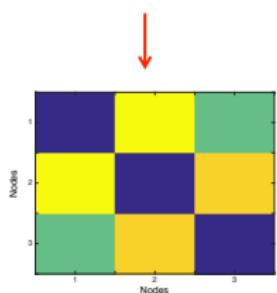


Induce structure on weights

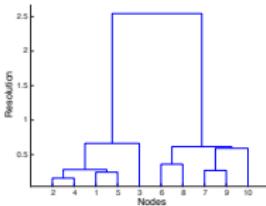


Weighted dissimilarity network

- Large dimension
- No structure on weights

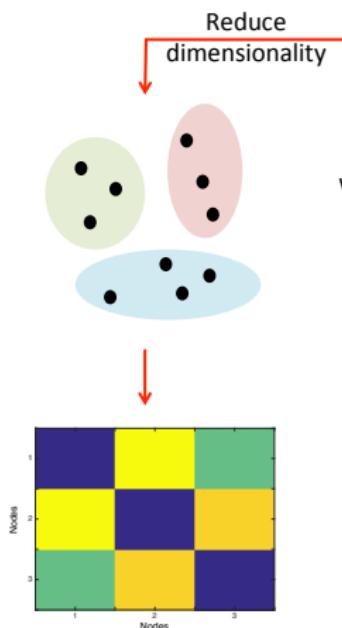


Smaller network

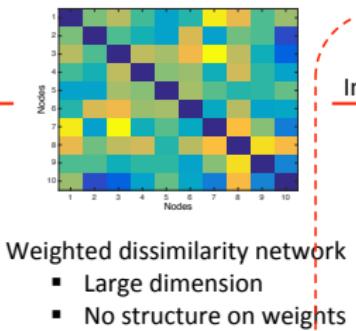


Structured representation

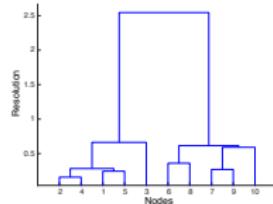
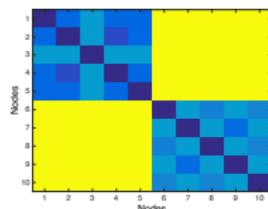
Clustering



Smaller network



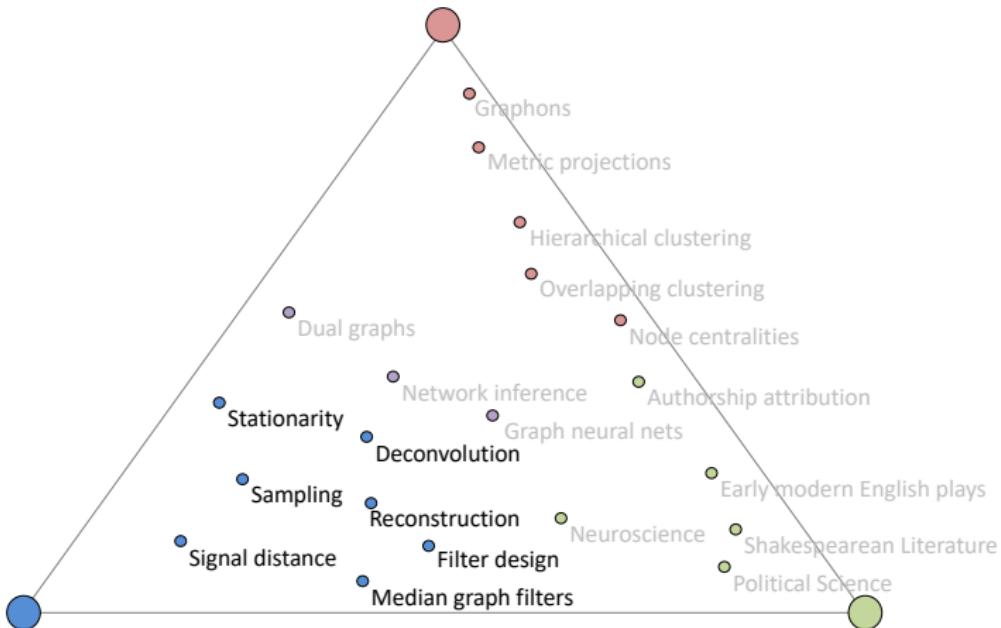
Induce structure on weights



Structured representation



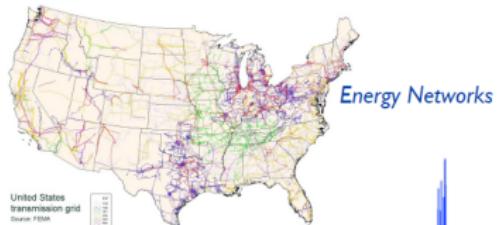
Understanding Networks



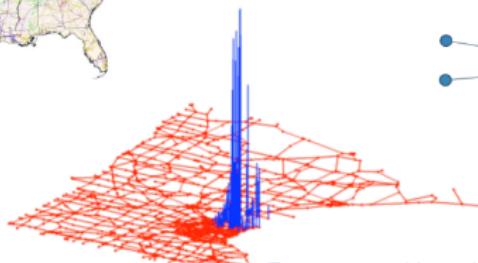
Understanding Network Data

Applications

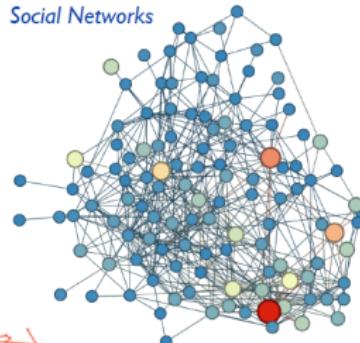
Motivating examples – Graph signals



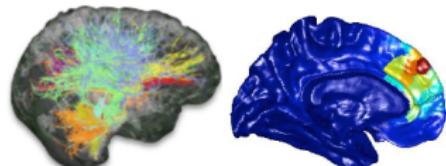
Energy Networks



Transportation Networks



Social Networks



Biological Networks



Irregular Data Domains

credit: LTS2

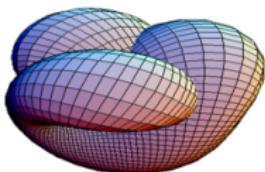
Motivating examples – Processing signals



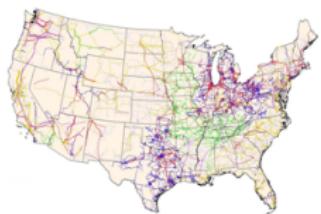
Interpolate a brain signal
from local observations



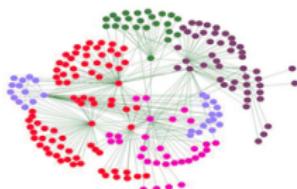
Compress a signal in
an irregular domain



Localize the
source of a rumor



Smooth an observed
network profile



Predict the evolution of a
network process

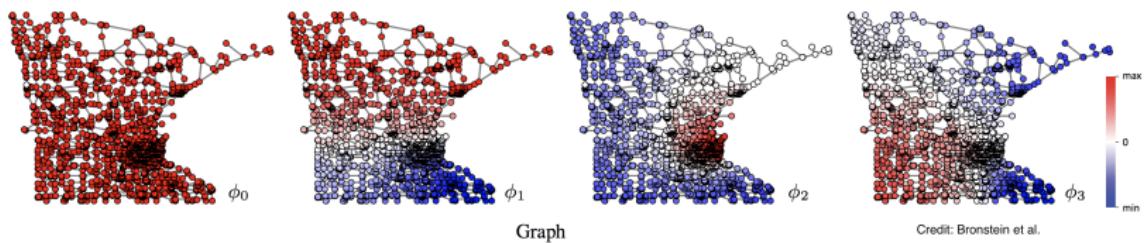
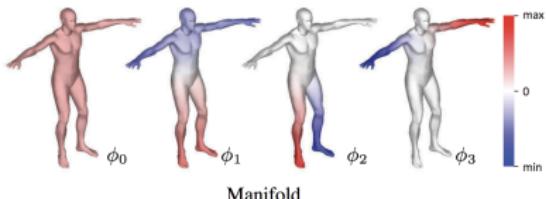
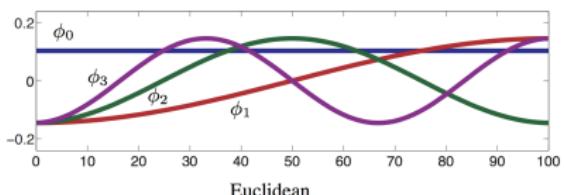


Infer the topology where
the signals reside

Fourier basis and beyond



- Generalize the notion of Fourier basis to non-Euclidean domains



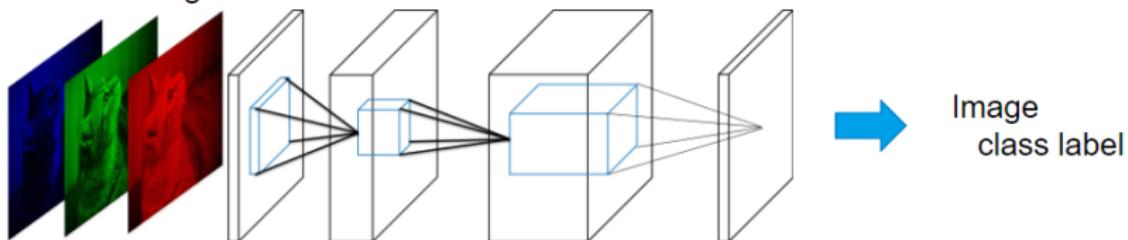
Credit: Bronstein et al.

- Notion of filtering \Rightarrow Smoothing, interpolation, compression
- System identification and (blind) deconvolution
- Stationary processes and power spectral estimation

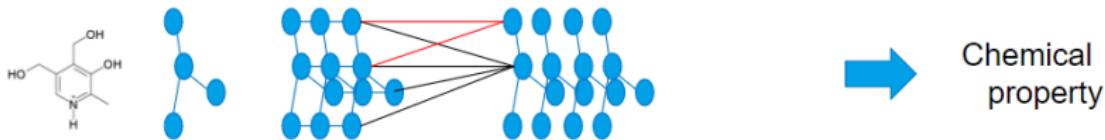
Graph convolutional neural networks



CNN on image



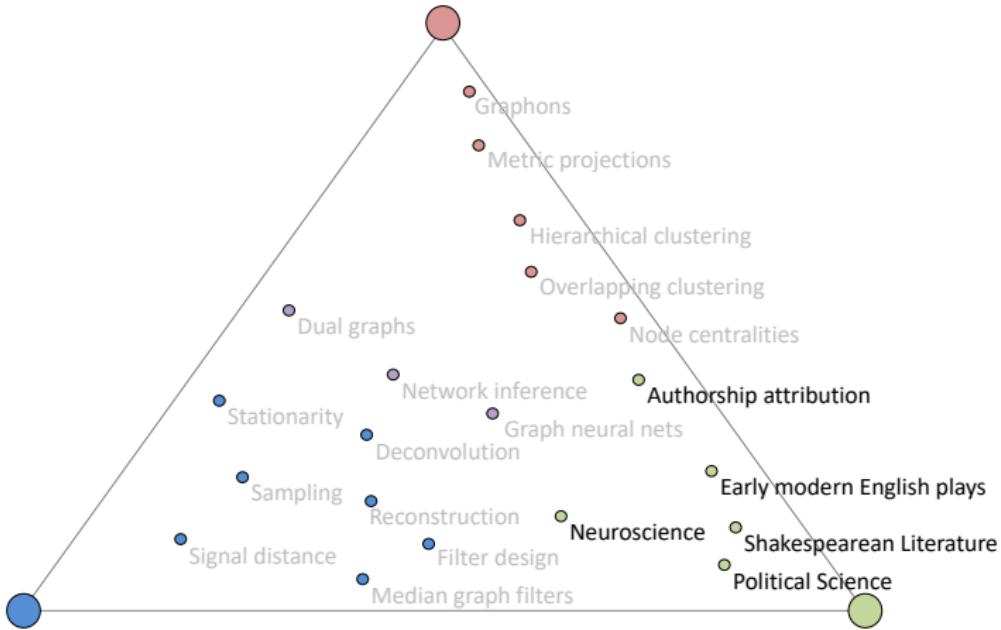
Graph convolution



Credit: Nakago



Understanding Networks



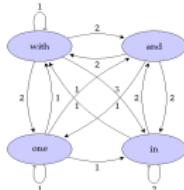
Understanding Network Data

Applications

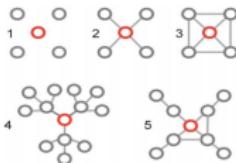
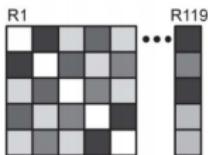
Motivating interdisciplinary questions



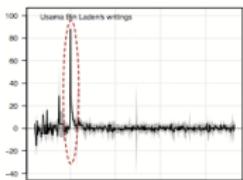
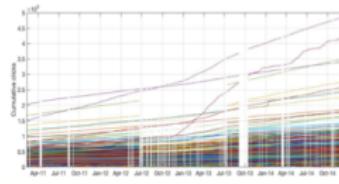
- ▶ Who *really* was William Shakespeare?



- ▶ How to diagnose Alzheimer's disease?



- ▶ What makes a *jihadist text* popular?



Shakespeare and the networks



Course logistics

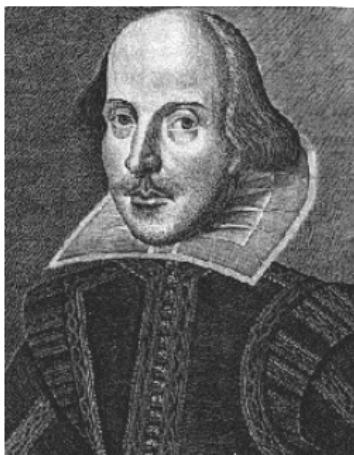
Networks: What are they and why are they important?

Data Science for Networks: A bit of what we do

To link or not to link: Shakespeare and the networks



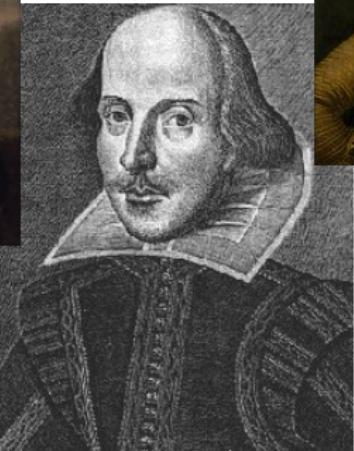
The authorship controversy



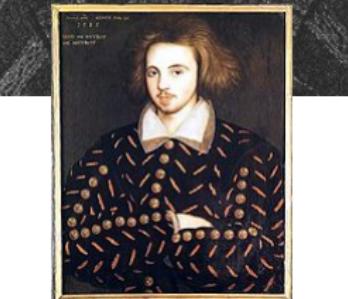
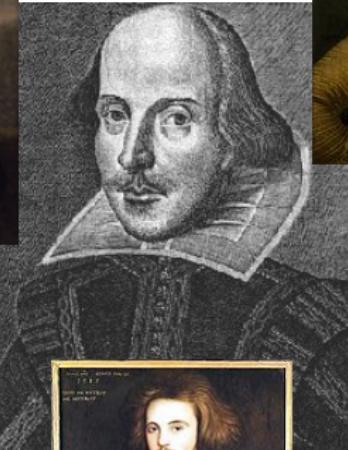
The authorship controversy



The authorship controversy



The authorship controversy

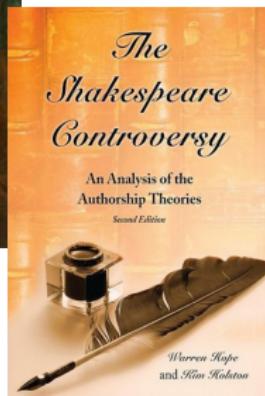
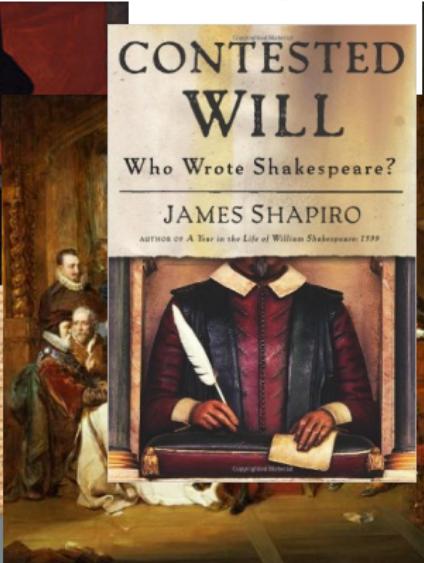
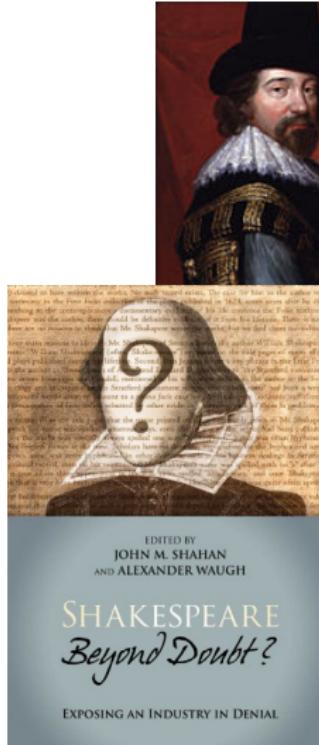




The authorship controversy



The authorship controversy

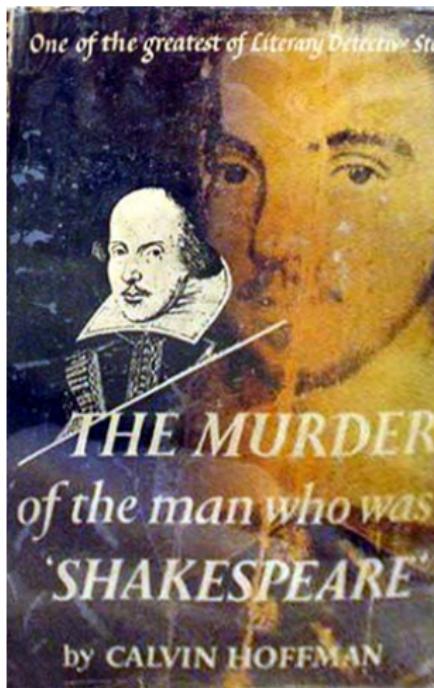




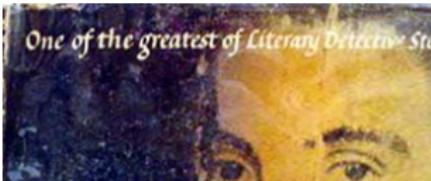
The authorship controversy



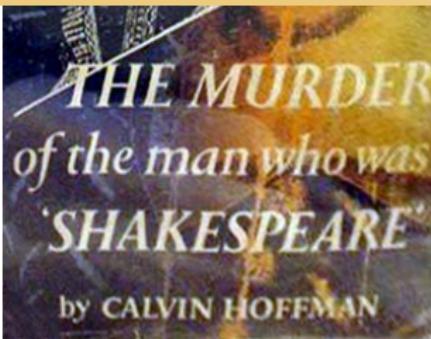
The Hoffman price



The Hoffman price

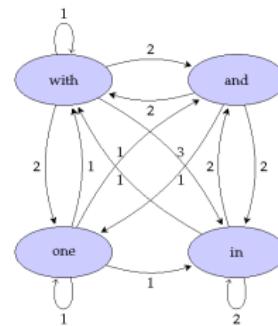


The principal prize will be awarded to the person who "has in the opinion of the King's School furnished irrefutable and incontrovertible proof and evidence required to satisfy the world of Shakespearian scholarship that all the plays and poems now commonly attributed to William Shakespeare were in fact written by Christopher Marlowe."





Word Adjacency Networks



- ▶ Authorship attribution tool based on the **redundancy** of the language
- ▶ Mathematical concept of Shannon **entropy**

Building WANs: An Example

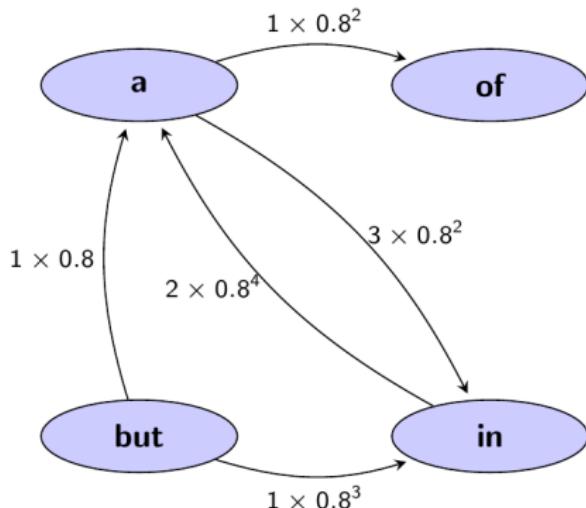


"A swarm in May is worth a load of hay; a swarm in June is worth a silver spoon; but a swarm in July is not worth a fly."

a swarm **in** May is worth a load **of** hay

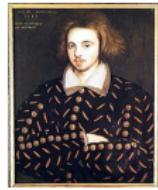
a swarm **in** June is worth a silver spoon

but a swarm **in** July is not worth a fly

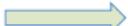
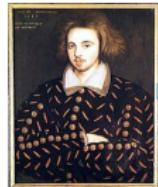


- ▶ An edge is the sum of mutual occurrences in text
- ▶ Each occurrence is discounted by a factor of 0.8 for every word separating them

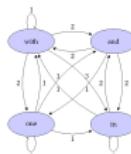
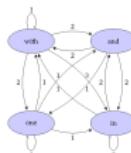
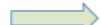
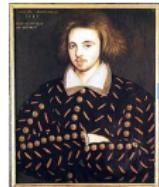
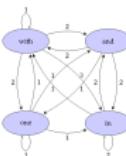
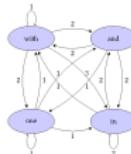
How does our method work?



How does our method work?

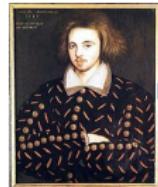
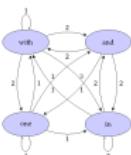
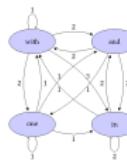


How does our method work?

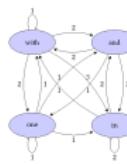
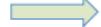
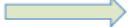
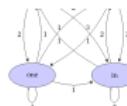




How does our method work?

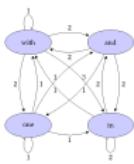
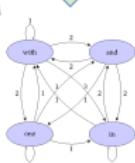
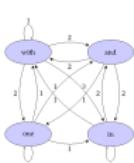
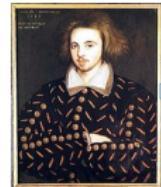
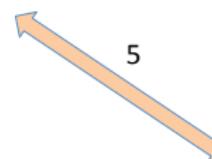
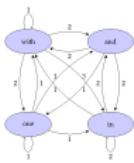


$$H(P_1, P_2) = \sum_{i,j} \pi(f_i) P_1(f_i, f_j) \log \frac{P_1(f_i, f_j)}{P_2(f_i, f_j)}$$



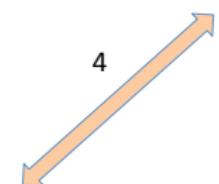
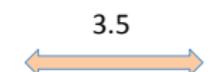
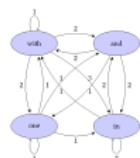
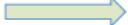
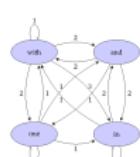
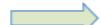
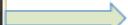
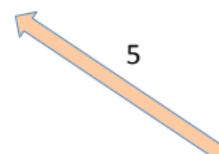
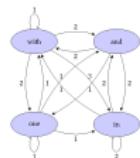


How does our method work?



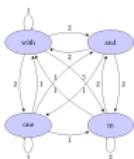
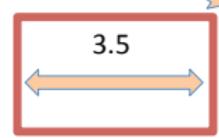
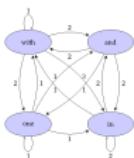
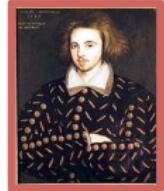
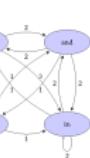
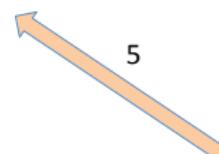
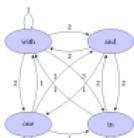
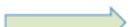


How does our method work?





How does our method work?





It works pretty well

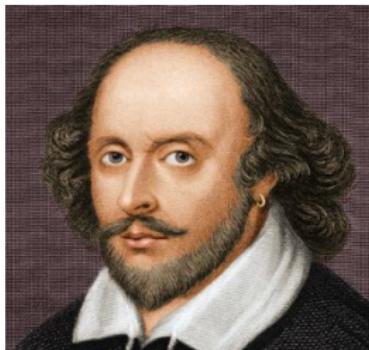
- ▶ Accuracy depends on a number of factors
 - ⇒ Number of **candidate authors**
 - ⇒ Length of the **training corpus**
 - ⇒ Length of the **text to attribute**
- ▶ Attribution of a text of 25,000 words
 - ⇒ Around the length of a Shakespearean play

Nr. of authors	Number of words in profile (thousands)										Rand.
	10	20	30	40	50	60	70	80	90	100	
2	0.927	0.964	0.984	0.985	0.981	0.979	0.981	0.986	0.992	0.988	0.500
3	0.871	0.934	0.949	0.962	0.968	0.975	0.982	0.978	0.974	0.978	0.333
4	0.833	0.905	0.931	0.949	0.948	0.964	0.963	0.968	0.969	0.977	0.250
5	0.800	0.887	0.923	0.950	0.945	0.951	0.953	0.961	0.961	0.969	0.200
6	0.760	0.880	0.929	0.932	0.937	0.941	0.948	0.952	0.950	0.973	0.167
7	0.755	0.851	0.909	0.924	0.937	0.943	0.937	0.957	0.960	0.957	0.143
8	0.722	0.841	0.898	0.911	0.932	0.941	0.938	0.947	0.952	0.955	0.125
9	0.711	0.855	0.882	0.905	0.915	0.931	0.932	0.944	0.952	0.955	0.111
10	0.701	0.827	0.882	0.910	0.923	0.923	0.934	0.935	0.943	0.935	0.100

Shakespearean debate



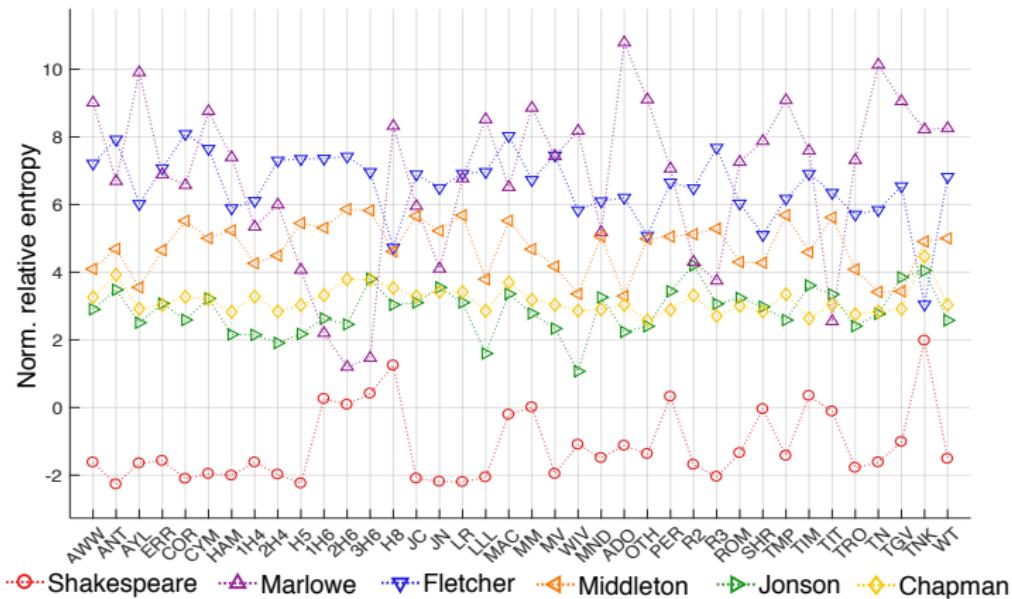
- ▶ **Big Question:** Did William Shakespeare have collaborators in his plays?
- ▶ Collaboration not well documented in late 1500s, early 1600s
 - ⇒ But scholars have identified possible candidates





Henry VI: a collaborative play?

- Level of collaboration of Shakespeare is still debated
- All of Shakespeare's plays correctly attributed to him
- Attribution of *Henry VI* plays \Rightarrow Supports a side of the debate





Impact on accepted authorship

New Oxford Shakespeare Edition Credits Christopher Marlowe as a Co-author

By CHRISTOPHER D. SHEA OCT. 24, 2016



A portrait of Christopher Marlowe, a Shakespeare colleague and rival. Corpus Christi College

LONDON — Shakespeare may have had a little more help than previously suspected.

The New Oxford Shakespeare edition of the playwright's works — which will be published by [Oxford University Press](#) online ahead of a worldwide print release — lists Christopher Marlowe as Shakespeare's co-author on the three "Henry VI" plays, parts 1, 2 and 3.

It's the first time that a major edition of Shakespeare's works has listed Shakespeare's colleague and rival as a co-author on these works, the volume's general editor, Gary Taylor, said in a phone interview.

Impact on accepted authorship



New Oxford Shakespeare Edition Credits Christopher Marlowe as a Co-author

By CHRISTOPHER D. SHEA OCT. 24, 2016



A portrait of Christopher Marlowe, a Shakespeare colleague and rival. Corpus Christi College

LONDON — Shakesp little more help than I suspected.

The New Oxford Shal the playwright's work published by [Oxford](#) online ahead of a wor — lists Christopher M Shakespeare's co-autl "Henry VI" plays, par

It's the first time that Shakespeare's works. Shakespeare's collega author on these work general editor, Gary T phone interview.

Oxford to make Christopher Marlowe a Shakespeare co-writer for Henry VI



At left, a painting of William Shakespeare which is believed to be the only authentic image of Shakespeare made during his life unveiled by The Shakespeare Birthplace Trust on March 9, 2009 in London, England. At right, what is believed to be a portrait of Christopher Marlowe, painted in 1585, and discovered by his alma mater in England, Corpus Christi College, Cambridge, in 1952. / THE SHAKESPEARE BIRTHPLACE TRUST/CORPUS CHRISTI COLLEGE/CBS NEWS COMPOSITE



New Oxford Shakespeare Edition Credits Christopher Marlowe as a Co-author

By CHRISTOPHER D. SHEA OCT. 24, 2016



LONDON — Shakesp
little more help than I
suspected.

William Shakespeare's Henry VI co-written by Christopher Marlowe, 'big data' analysis reveals

Posted 25 Oct 2016, 12:20am

A new edition of William Shakespeare's complete works will name Christopher Marlowe as co-author of three plays, shedding new light on the links between the two great playwrights after centuries of speculation and conspiracy theories.

Marlowe will be listed as co-author of the three Henry VI plays in the New Oxford Shakespeare, due to be published in several instalments over the coming weeks by the Oxford University Press.

"Shakespeare has entered the world of big data and there are certain questions that we are now able to answer more confidently that people have been asking for a very long time," Gary Taylor, one of the project's senior editors, told Reuters.

The issue of whether Shakespeare wrote all the plays attributed to him has been the subject of endless conjecture, with one persistent theory being that they were actually written by Marlowe — a notion rejected by Shakespeare scholars.

Professor Taylor, a Florida State University expert, said academics had known for a long time that Shakespeare worked with other writers on some plays.

The idea that he collaborated with Marlowe on the

Oxford to make Christopher Marlowe a Shakespeare co-writer for Henry VI

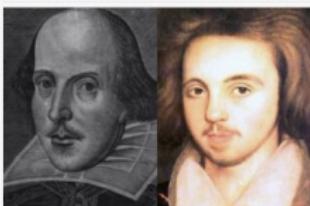


PHOTO: Christopher Marlowe (R) "very clearly" wrote parts of Shakespeare's Henry VI plays, experts said. (Supplied)

RELATED STORY: If Shakespeare were alive he'd be writing Game of Thrones'

RELATED STORY: Rare Romeo and Juliet book that inspired Shakespeare's play on display

MAP: England



which is believed to be the only authentic image of Shakespeare made eare Birthplace Trust on March 9, 2009 in London, England. At right, what is Marlowe, painted in 1585, and discovered by his alma mater in England, 1952. / THE SHAKESPEARE BIRTHPLACE TRUST/CORPUS CHRISTI

Key points:

- Marlowe's role in the Henry VI plays had been debated for centuries
- Academics extremely confident about Marlowe's authorship of some parts
- Collaboration between playwrights was normal

