

# **CSE 640:**

# **Graph Mining and Management**

## **Lecture 1 (Feb 2)**

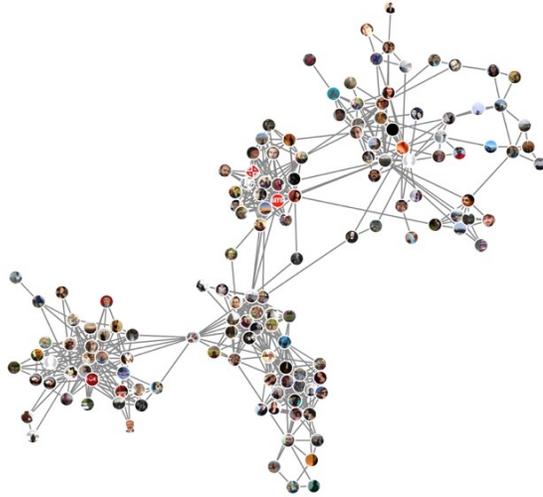
A. Erdem Sariyuce

# Introductions

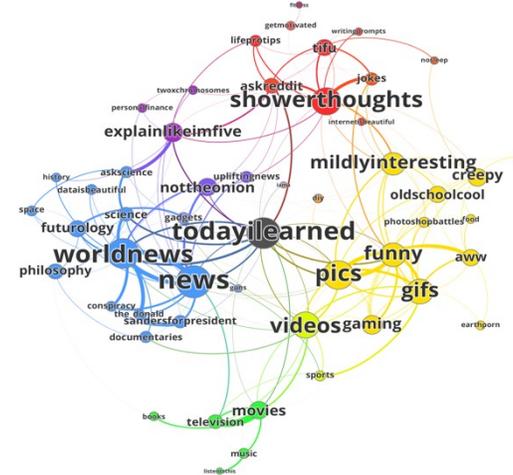
- My name is A. Erdem Sariyuçe
  - I go by **Erdem**
  - My pronouns are he/him/his
  - And I'd be happy if you call me Erdem
- Assistant Professor in CSE since 2017
  - My research is on graph mining
- I like NBA and binge-watching TV series
  - Appreciate any suggestion!
- You? Please mention your pronouns

# Graphs (networks) are everywhere

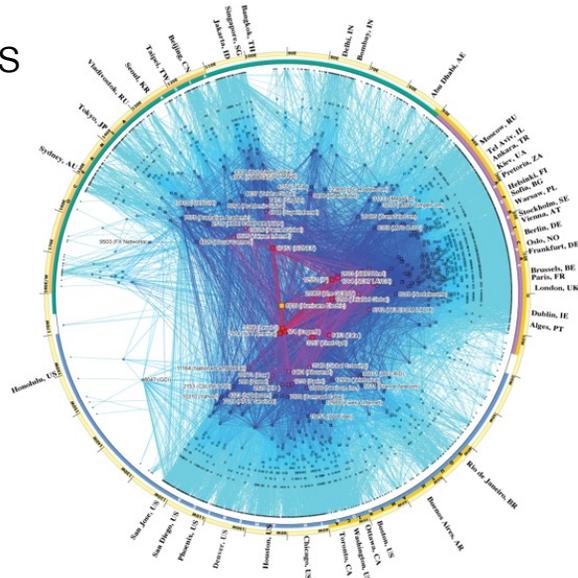
Social



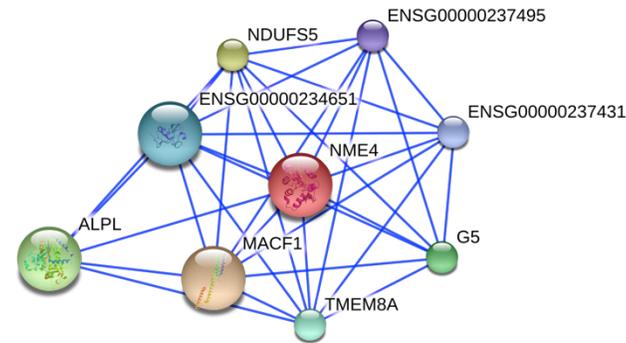
Information



Routers



Protein-interaction



# Transformers of disciplines

- Graph theory
  - Mathematics
- Statistical mechanics
  - Physics
- Data mining, ML
  - Computer science
- Inferential modeling
  - Statistics
- Social structure
  - Sociology



# Questions we ask

- Real-world networks
  - What **characteristics** do we observe?
- Nodes and edges
  - Which ones are **more important** than others?
  - What does important mean in this context?

# Questions we ask

- Given a person in a social network;
  - How do we determine her social circles?
  - How do we suggest new friends?
  - Can we infer her spouse? (there is a paper on that)
- Or given a webpage about booking flights,
  - How likely it will be accessed next week?
  - How can we make it appear on top in search results?

# Questions we ask

- Graph algorithms
  - What is used in the state-of-the-art [search engines](#)?
  - What about [recommendation systems](#)?
- How does a network [evolve over time](#)?
  - Which nodes will get more edges?
  - Which edges will be removed or added?

# Logistics

- Course website: <https://sariyuce.com/S22-640.html>
- Class hours: Mon, Wed 1:30-2:50 @ NSC 216
- Office hours: Wed 3:30-5:30 over Zoom (link will be provided)
  - No office hours today; post your questions on Piazza
- I prefer to do all communications over Piazza
  - Including private messages
- For private/urgent matter: erdem@buffalo.edu

# Logistics

- CSE 531 is a prerequisite
  - Background in graph theory, discrete math
  - Programming in some language
- No textbook required. But will benefit from
  - Networks: An Introduction
    - By M. Newman
  - Networks, Crowds and Markets
    - By D. Easley and J. Kleinberg
    - <https://www.cs.cornell.edu/home/kleinber/networks-book>

# Lectures and papers

- For advanced topics (second half of semester), we will mainly consider recent papers
  - Community Detection, Temporal Networks ...
- Great papers from the top venues!
  - Science, Nature
  - SIGKDD, WWW, WSDM, ICDM, SDM
  - VLDB, SIGMOD, ICDE

# Grading

- Homeworks
  - 30%
- Midterm
  - 20% (In class, date TBD)
- Project
  - 50%

# Schedule

- See course webpage
  - The entire semester will be posted
  
- Spring recess: Mar 21 – Mar 26

# Homeworks

- Combination of data analysis, algorithm design, and math
  - Analysis and discussions by charts, tables
  - Require light coding to automate things
- Three homeworks
  - Due in one week
  - Individually
- All homeworks should be typed!
  - Submission via AutoLab (see course webpage for more details)

# Project

- Proposal by 3<sup>rd</sup> week
  - Report
  - Short presentation
- Progress after Spring break
  - Report
  - Presentation
- Final by the last week
  - Report (May 9)
  - Presentation

- Can form teams of two
  - Don't have to but encouraged
- Ideas will be provided
  - Don't worry, I'll guide
- We have to meet every other week, at least
  - Office hours! Wed 3:30-5:30

# Academic integrity

- Don't cheat in homeworks!
- New university policy since Fall 2019!
  - <https://www.buffalo.edu/academic-integrity/about/process.html>
  - Any incident has to be reported to AI office and goes into student's record
- Zero tolerance: Failure in the course for first attempt
- Grads: Sanctions can even reach to RA/TA cancellation

# Accessibility Resources

- If you have a diagnosed disability (physical, learning, or psychological) that will make it difficult for you to carry out the course work as outlined, or that requires accommodations such as recruiting note-takers, readers, or extended time on exams or assignments, you must consult with [Accessibility Resources](#) (60 Capen Hall: 716-645-2608).
- You must advise me during the first two weeks of the course so that we may review possible arrangements for reasonable accommodations.
- (Also available in the course webpage)

# Critical Campus Resources

- **Sexual Violence**

- UB is committed to providing a safe learning environment free of all forms of discrimination and sexual harassment, including sexual assault, domestic and dating violence and stalking.
- If you have experienced gender-based violence (intimate partner violence, attempted or completed sexual assault, harassment, coercion, stalking, etc.), UB has resources to help.
- This includes academic accommodations, health and counseling services, housing accommodations, helping with legal protective orders, and assistance with reporting the incident to police or other UB officials if you so choose. Please contact UB's Title IX Coordinator at 716-645-2266 for more information. For confidential assistance, you may also contact a Crisis Services Campus Advocate at 716-796-4399.

- **Mental Health**

- [Counseling Services](#)

- 120 Richmond Quad (North Campus), 716-645-2720
- 202 Michael Hall (South Campus), 716-829-5800

- [Health Services](#)

- Michael Hall (South Campus), 716-829-3316

- [Health Promotion](#)

- 114 Student Union (North Campus), 716-645-2837

# Preferred Name

- If you would like to be addressed by a name that is different from the one in UB records, please let me know and we will use your preferred name in our communications with you.
- Further, you will be able to use your preferred name in all of your exams, homeworks, and project-related documents.

**Questions?**

# Project Ideas

- 1) **Surveys** on certain hot topics
  - Surveys are key for advancing research
  - With a codebase for comparison
- 2) **Specific topics** that I'm doing research on
  - I have very specific tasks and execution plans
  - More on this later on
- 3) **Repeatability experiments** for some popular papers
  - And extensions
- 4) **Data Cup (or ML)** challenges
- 5) **Any idea** you may want to go for!
  - Consultation with instructor
- Promising projects will be continued after the semester! Funded positions available.

# 1) Surveys

- Surveys are the most cited type of research papers
  - Helps advancing the field in a rigorous way
  - Textbook is the ultimate form of knowledge
- Based on hot research topics
  - Start with a particular paper X
  - Check the papers that cited X via Google Scholar
    - And continue like BFS!
  - Read the abstracts and intros in those papers
    - Read more only if it's directly related
- Objective experimental comparison
  - You become the judge and provide a holistic evaluation
  - Explore the parameter spaces; reveal the hidden assumptions; try out more datasets
- Best for groups; might be too much for a single person

## 2) Specific topics

- Temporal analysis of financial transaction networks
  - Cryptocurrency transactions
  - Anomaly detection
- Fair graph mining
  - Protected-label based bias-avoiding community detection
- Signed network motifs
  - Building on a project in last year's class
- Drug repurposing
  - Using a bipartite drug-disease network; COVID-19 related measurements
- Core-periphery structure using network motifs
  - By utilizing an existing framework

# 3) Repeatability experiments

- Science has a repeatability problem
  - [https://en.wikipedia.org/wiki/Replication\\_crisis](https://en.wikipedia.org/wiki/Replication_crisis)
  - A 2019 study reporting a systematic analysis of recent publications applying deep learning or neural methods to recommender systems, published in top conferences (SIGIR, KDD, WWW, RecSys)
    - Less than 40% of articles are reproducible, with as high as 75% and as little as 14% depending on the conferences.
    - All (but one) of the algorithms were not competitive against much older and simpler properly tuned baselines.
    - <https://dl.acm.org/doi/10.1145/3298689.3347058>
- Graph mining and network science are no different
  - Graph dependent approaches with limited datasets
  - Making up nonsense metrics that are hiddenly aimed by the proposed algorithm
  - And more
- Choose a related paper published in a top venue recently and replicate the results. And extend further!
  - What lessons learned? Which experiments are biased? What happens when used on some other datasets?

# 4) Data Cup (ML) Challenges

- Some conferences have challenges for competition
  - A well-defined task
  - With specific input/output and metric
- If the challenge fits well to 640, we can go for it
- Watch Piazza for more
  - I'll post the details
- Only if you have time and energy

# 5) Any idea you may want to go for!

- Talk to me!
- What you have in mind
  - Might not be related to our focus in this class
  - Might not worth a semester-long project
  - Might be too ambiguous for a semester-long project
- You don't have to have a very well-defined thing, vague directions are OK too
  - But you must talk to me

# Proposal deadline: Feb 16<sup>th</sup> 1:30pm

- You should
  - determine your groups,
  - think about your topic,
  - discuss it with me,
  - and finalize the report and presentation by answering the questions below.
    - What is the problem?
    - Why do we care?
    - What is your execution plan?
- Start today!!
- By Feb 16<sup>th</sup>, Wed: Report is due 1:30pm, Presentation is in-class

# Questions?