Graphs and Complex Networks Across Domains

GraphXD Workshop 2018

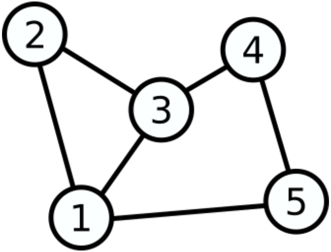
Berkeley Institute for Data Science (BIDS)

Jarrod Millman
Division of Biostatistics
University of California, Berkeley

March 27, 2018

What is a graph?

A set of **vertices** connected by **edges**

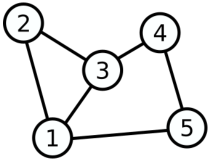


Formally, the graph G is an ordered pair (V, E) where

$$V = \{1, 2, 3, 4, 5\}$$
$$E = \{\{1, 2\}, \{1, 3\}, \{1, 5\}, \{2, 3\}, \{3, 4\}, \{4, 5\}\}$$

Matrix perspective

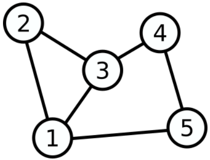
Adjacency matrix



$$A_G = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$V = \{1, 2, 3, 4, 5\}$$
$$E = \{\{1, 2\}, \{1, 3\}, \{1, 5\}, \{2, 3\}, \{3, 4\}, \{4, 5\}\}$$

Laplacian = Degree − Adjacency



$$L_G = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$V = \{1, 2, 3, 4, 5\}$$
$$E = \{\{1, 2\}, \{1, 3\}, \{1, 5\}, \{2, 3\}, \{3, 4\}, \{4, 5\}\}$$

Linear Algebra

Spectral Theorem for Symmetric Matrices

Definition

A scalar λ is called an **eigenvalue** of a square matrix A and a vector v its associated **eigenvector** if $Av = \lambda v$.

Theorem

Let A be a symmetric $n \times n$ matrix with real-valued entries, then there are n (not necessarily distinct) real eigenvalues

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \qquad \text{(spectrum)}$$

and n orthonormal real vectors

$$v_1, v_2, \ldots, v_n$$

such that v_i is an eigenvector of λ_i.

Where do graphs come from?

Graphs from data

Vertices

- People
- Animals
- Ancestors
- Cities
- Genes
- Brain regions
- Random variables
- Etc.

Edges

- Interactions
- Cooccurrence
- Expert knowledge
- Correlation
- Conditional independence
- Causality
- Etc.

Example

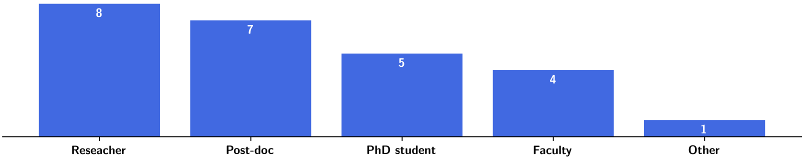
Pre-workshop Questionnaire

- Vertices (25 respondents out of 30)
  - Attributes
    - What best describes you?
    - I work in ...
    - How often do you?
- Edges
  - Observe
    - Collaborate / Communicate / Familiar with their work / Don't know them
    - Use their software / data / algorithms
  - Infer

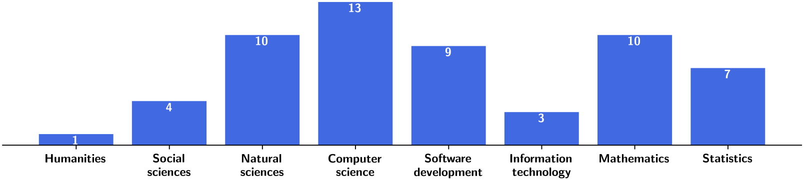
Vertices

Who are we?

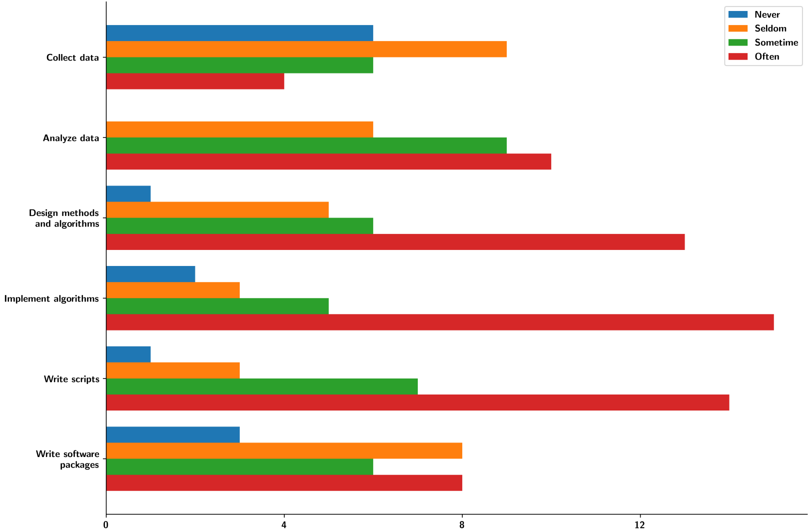
What best describes you?



I work in ...

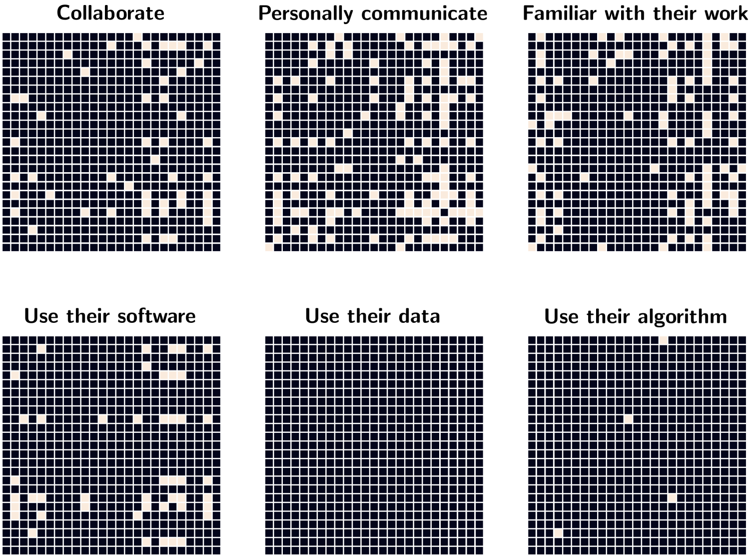


How often do you?

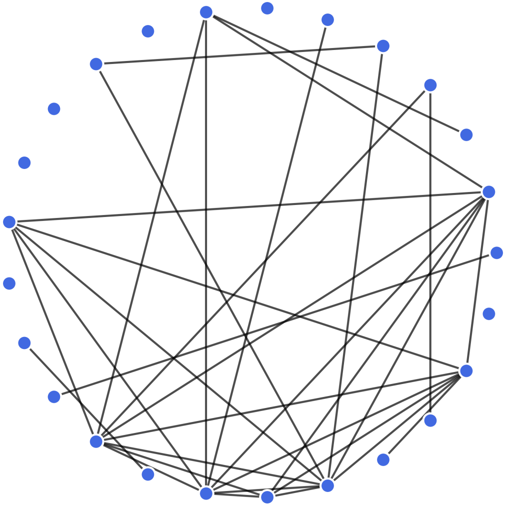


Edges

Adjacency matrices (respondent × participant)



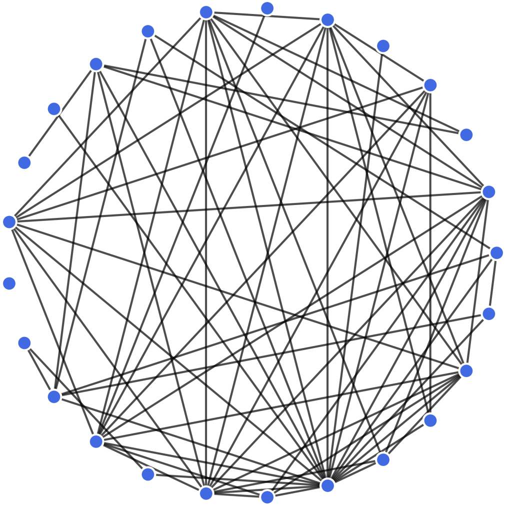
Collaborate ($n = 25, m = 33$)



Density: 0.11

Average degree: 2.6

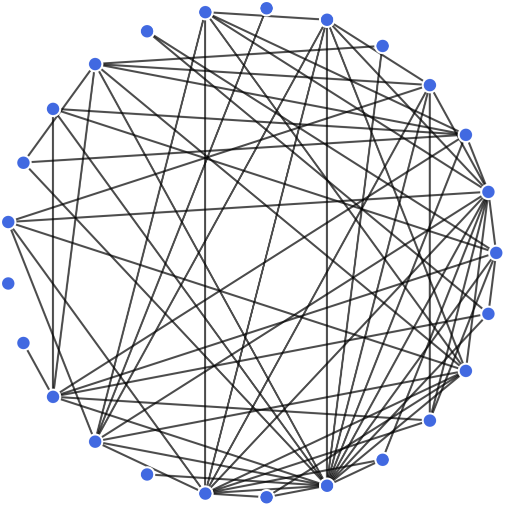
Personally communicate ($n = 25, m = 68$)



Density: 0.23

Average degree: 5.4

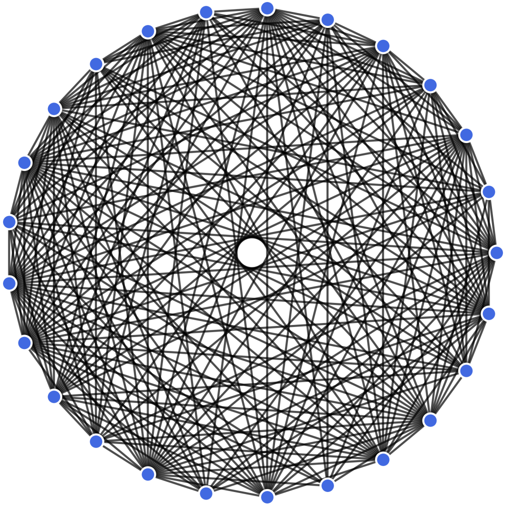
Familiar with their work ($n = 25, m = 70$)



Density: 0.23

Average degree: 5.6

Don't know them ($n = 25, m = 232$)

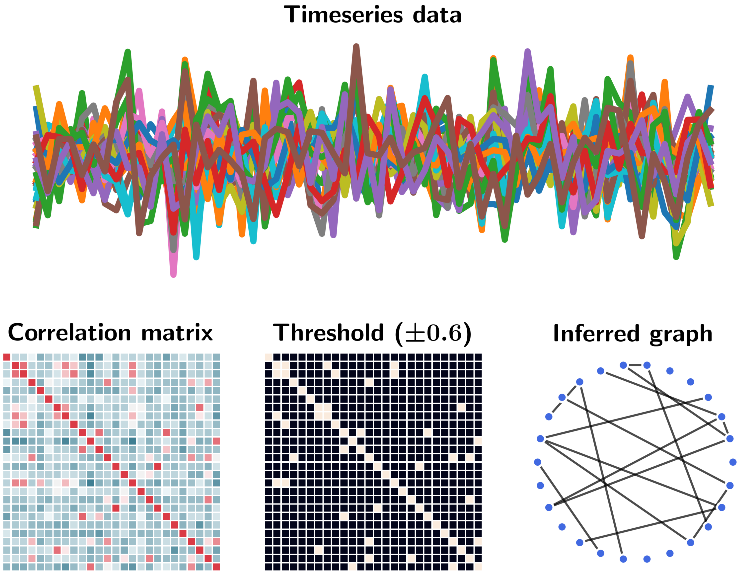


Density: 0.77

Average degree: 18.6

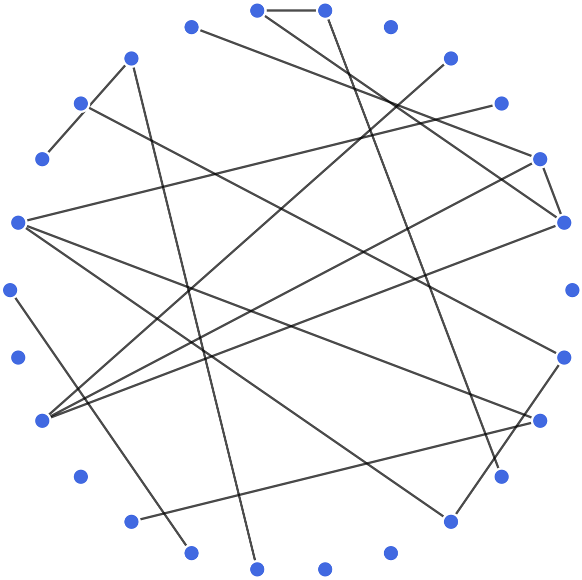
Inferring edges

Thresholded correlation

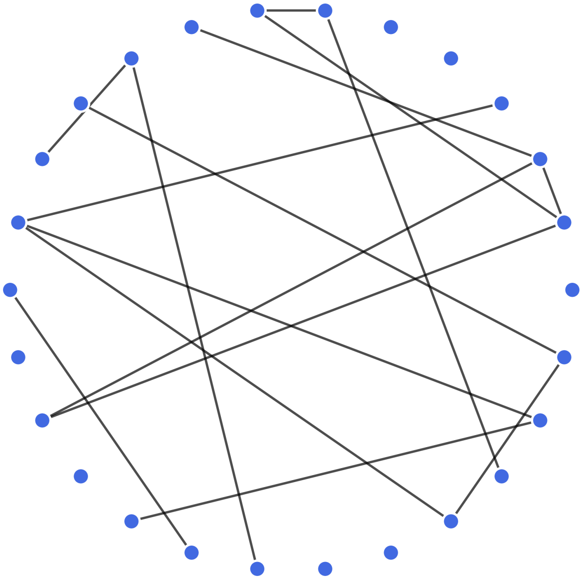


How well did we do?

Actual



Inferred



Plan

- Talks
- Food
- Discussion
- Self-organized activities
  - Scientific / research
  - Technical (e.g., coding sprints, tutorials, data analysis)
  - Communication (e.g., proposals, whitepapers, blog posts)
- Reflection

Thanks

- Stacey Dorton
- Jessica Zosa Forde
- Aaron Schild
- Stéfan van der Walt
- Nelle Varoquaux
- and the whole BIDS community ...

Discussion

Histogram of correlation coefficients