Èdouard Lucas:

The theory of recurrent sequences is an inexhaustible mine which contains all the properties of numbers; by calculating the successive terms of such sequences, decomposing them into their prime factors and seeking out by experimentation the laws of appearance and reproduction of the prime numbers, one can advance in a systematic manner the study of the properties of numbers and their application to all branches of mathematics.

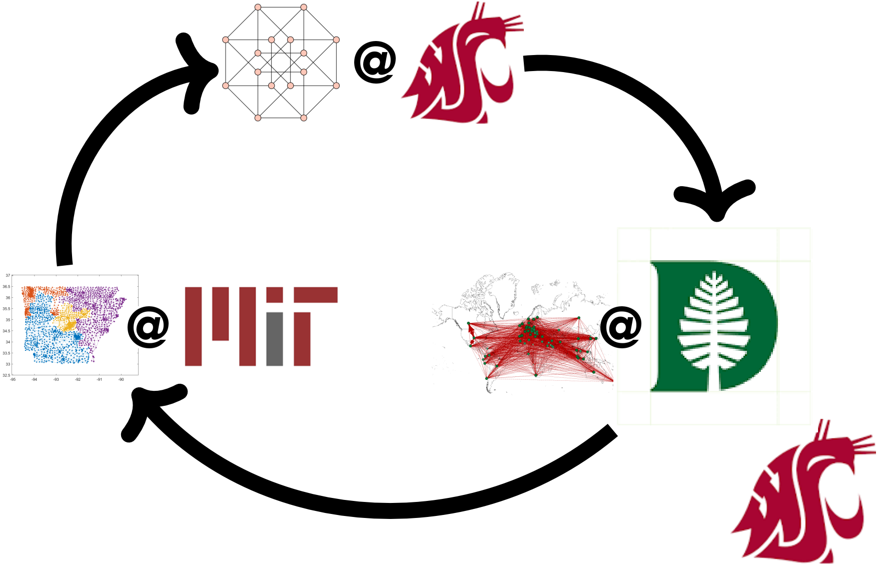
Social Network Analysis in Python

Daryl DeFord

Washington State University
Department of Mathematics and Statistics

CISER Workshop
March 2, 2023

Personal Background



Today's Outline

- **Goals:**
  - Try out some social network modeling!
  - Networks Networking
- **Format:**
  - Brief overview of ideas
  - Walk through notebooks
  - Highlight good places to experiment, change the parameters, extensions
- **Topics:**
  - Social network modeling
  - Introduction to Python
  - Jupyter Notebooks
  - Small Networks and Basic Commands
  - Plotting Networks
  - Social Network Properties
  - Dynamics on Networks

Interdisciplinary Techniques

- Communication
- Mathematics
- Statistics
- Sociology
- Teaching and Learning
- PPPA
- Criminal Justice
- Prevention Science
- Hospitality and Business Management
- Finance
- Education
- Anthropology
- Psychology
- ...

Additional Resources

- Workshop Github Page: https://github.com/drdeford/CISER2023NetworkX
- Notebooks on GitHub
- Links inside notebooks
- NetworkX Documentation
- Math 581 Webpage
- **Data 302**
- Data Sources:
  - SNAP
  - Newman's Networks
  - UCI Network Repository

Networked Otters

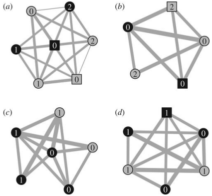


Figure from: Social Learning in Otters, Z. Ladds, W. Hoppitt, and N. J. Boogert, Royal Society of Open Science, 4(8), 2017.

Warmup Problem

Definition (Ego Network)

An ego network is a social network centered at a particular individual containing their connections and the connections between their "friends."

Warmup Problem

Definition (Ego Network)

An ego network is a social network centered at a particular individual containing their connections and the connections between their "friends."

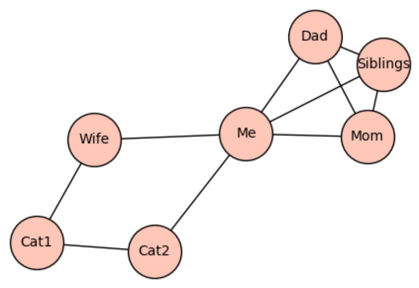
Problem (Draw your ego network)

Warmup Problem

Definition (Ego Network)

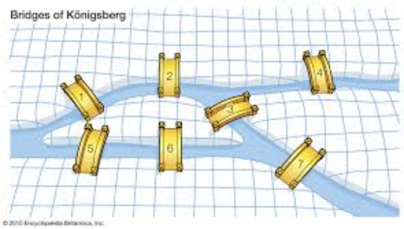
An ego network is a social network centered at a particular individual containing their connections and the connections between their “friends.”

Problem (Draw your ego network)



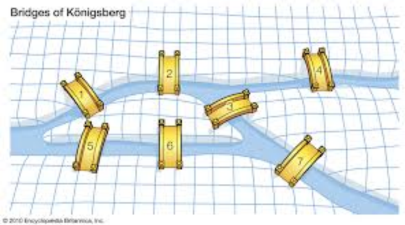
Historical Overview

- Euler (bridges of Königsberg)



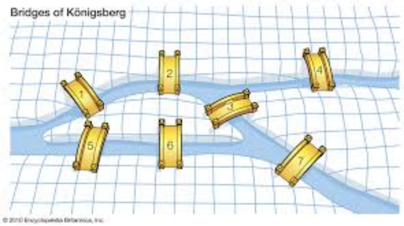
Historical Overview

- Euler (bridges of Königsberg)
- Sociologists
- Physical Infrastructure



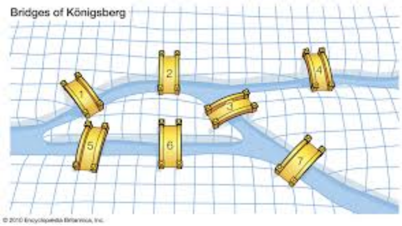
Historical Overview

- Euler (bridges of Königsberg)
- Sociologists
- Physical Infrastructure
- The internet!



Historical Overview

- Euler (bridges of Königsberg)
- Sociologists
- Physical Infrastructure
- The internet!



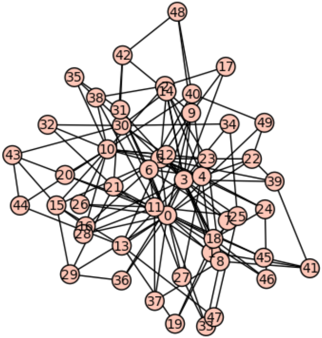
- Directed
- Weighted
- Hypergraphs
- Multiplex

Historical Overview

- Euler (bridges of Königsberg)
- Sociologists
- Physical Infrastructure
- The internet!

- Directed
- Weighted
- Hypergraphs
- Multiplex

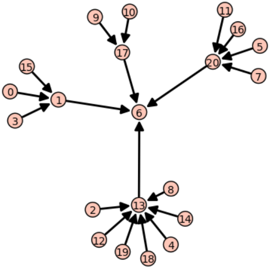
Network Examples



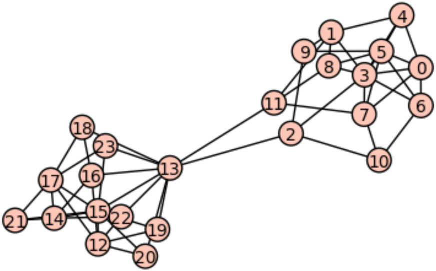
Network Examples

- Internet
  - Nodes: Hardware
  - Edges: Physical Connections
- WWW
  - Nodes: Webpages
  - Edges: Links
- Neuroscience
  - Nodes: Brain areas
  - Edges: Functional Connections
- Food webs
  - Nodes: Species
  - Edges: Predation
- World Trade
  - Nodes: Countries
  - Edges: Trade Agreements
- Banking
  - Nodes: Banks
  - Edges: Loans
- Genetics
  - Nodes: Genes/Proteins
  - Edges: Functional Interactions
- Recommendations
  - Nodes: People/Products
  - Edges: Ratings/Consumption

Network Questions



Centrality



Clustering

Centrality

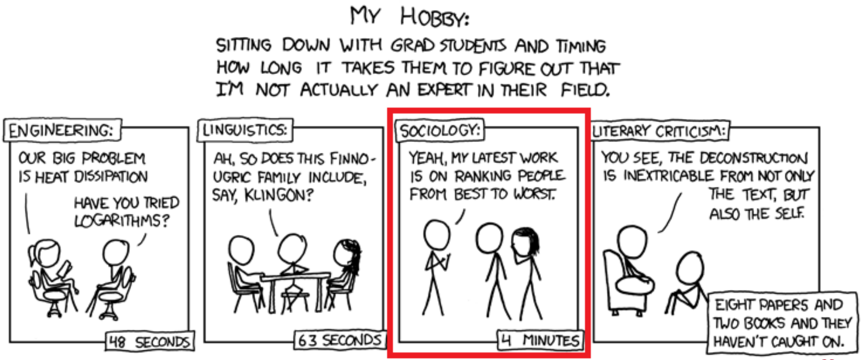
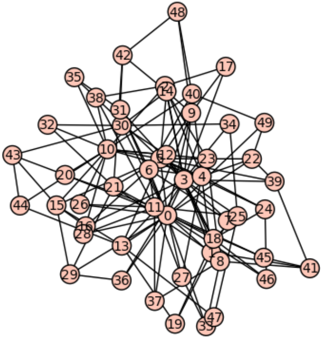


Figure: Relevant comic by Randall Munroe[1] (emphasis mine).

[1] https://xkcd.com/451/

Harder Questions



What is a social network?

Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

What is a social network?

Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

Example

What kinds of systems can social networks describe?

- What could be represented by nodes?

What is a social network?

Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

Example

What kinds of systems can social networks describe?

- What could be represented by nodes?
  - **WSU Students**

What is a social network?

Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

Example

What kinds of systems can social networks describe?

- What could be represented by nodes?
    - **WSU Students**
- What type of edges could connect them?

What is a social network?

Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

Example

What kinds of systems can social networks describe?

- What could be represented by nodes?
  - **WSU Students**
- What type of edges could connect them?
  - **In a class together**
  - **Facebook friends**
  - **Speak at least twice a week**

What is a social network?

Definition (Social Network)

Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

Example

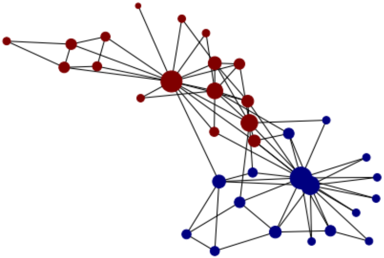
What kinds of systems can social networks describe?

- What could be represented by nodes?
  - **Academic Departments**
- What type of edges could connect them?
  - **Located in same building**
  - **Students who major in both**
  - **Crosslisted courses**

Common Properties of Social Networks

Example (What features distinguish social networks?)

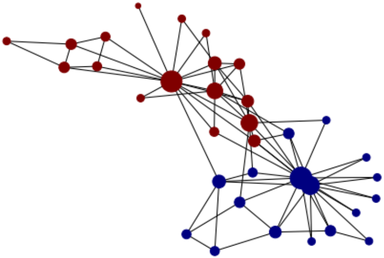
- Transitivity
- Community structure
- Small average path length
- Long–tailed degree distribution
- Hubs
- ...



Common Properties of Social Networks

Example (What features distinguish social networks?)

- **Transitivity**
- Community structure
- **Small average path length**
- Long–tailed degree distribution
- Hubs
- ...



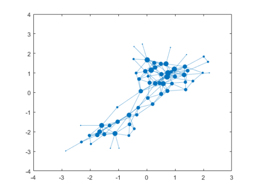
How to construct networks?

Example (Which edges to add?)

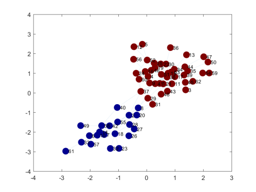
How to construct networks?

Example (Which edges to add?)

- Proximity



Centrality

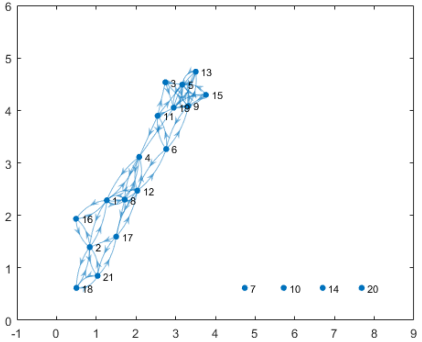


Clustering

Figure: Dolphin social network[1]

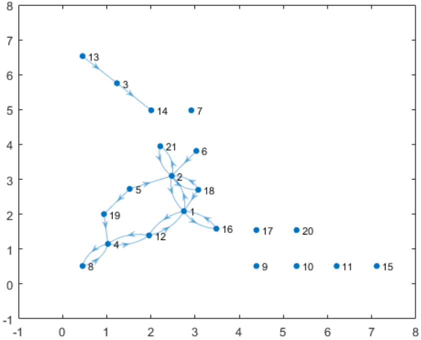
[1] D. Lusseau, K. Schneider, O. Boisseau, Patti Haase, E. Slooten, and S. Dawson, The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations, Behavioral Ecology and Sociobiology 54 (2003), 4, 396–405.

Different Perspectives on “Friendship”



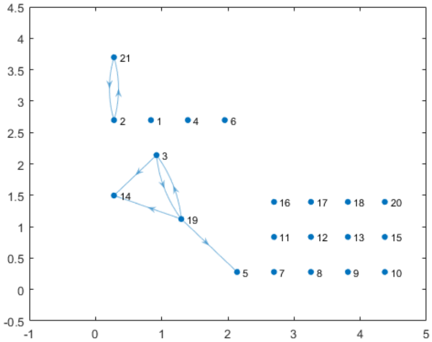
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



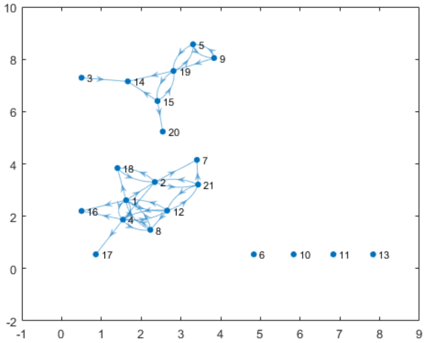
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



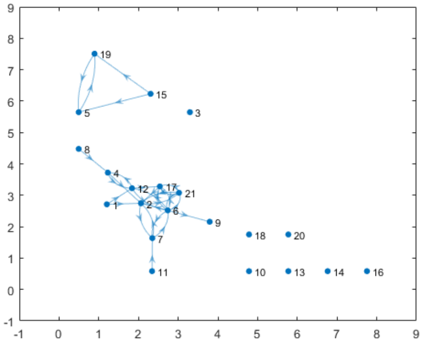
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



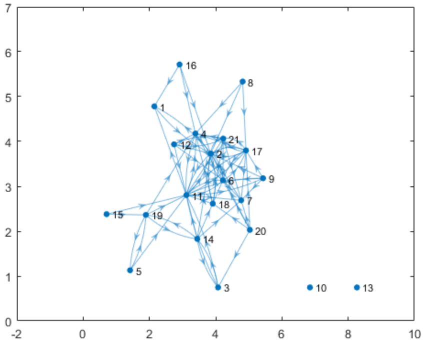
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



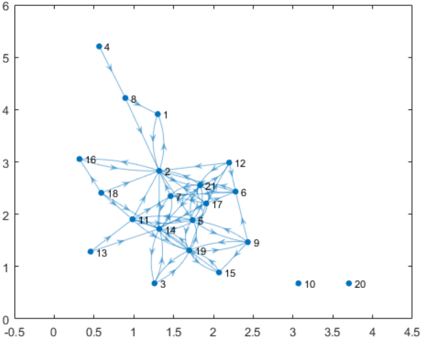
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



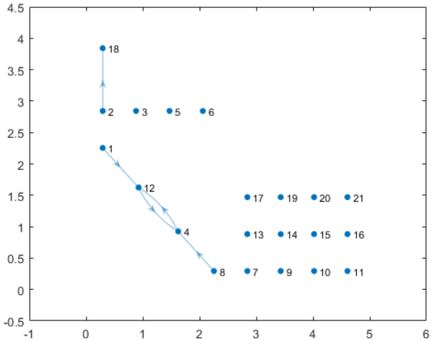
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



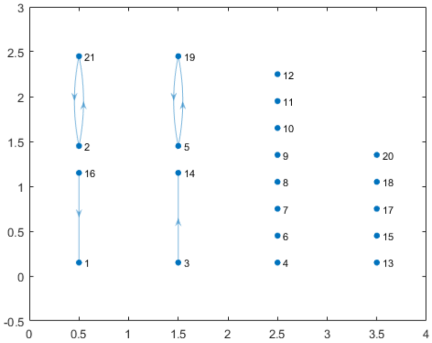
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



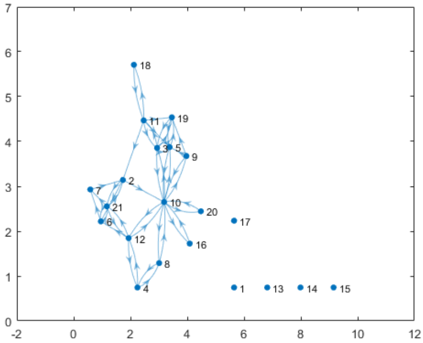
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



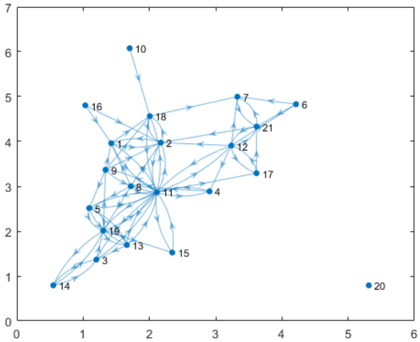
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



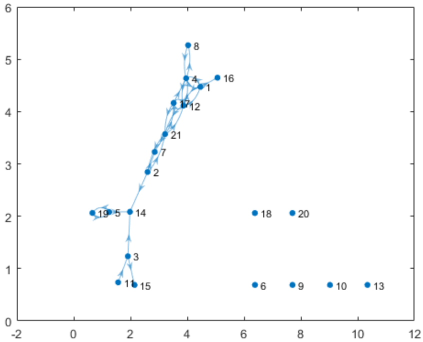
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



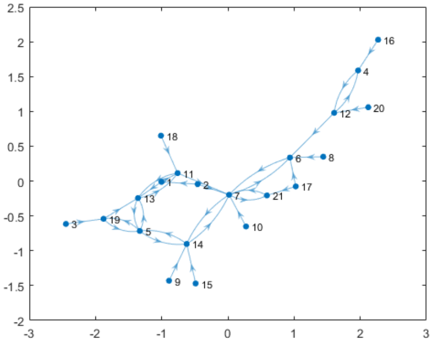
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



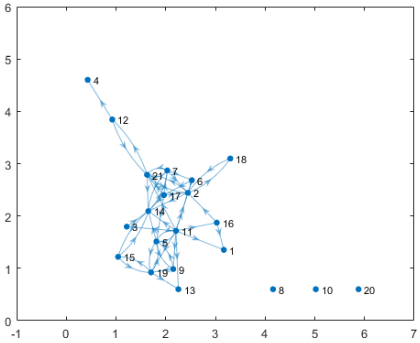
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



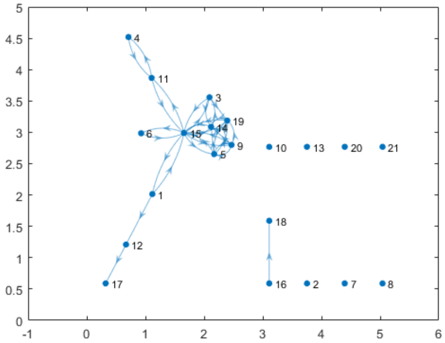
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



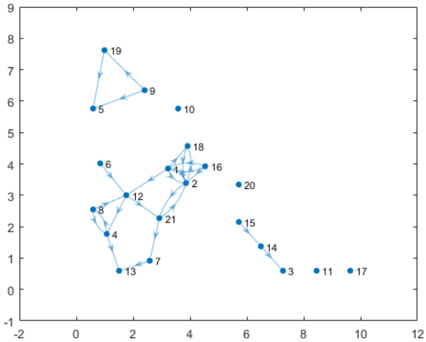
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



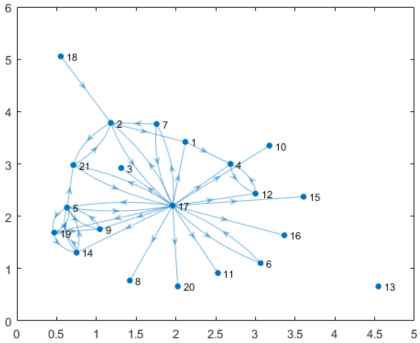
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



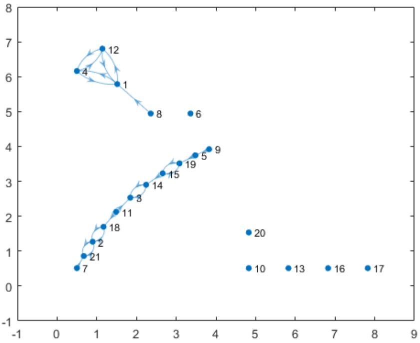
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



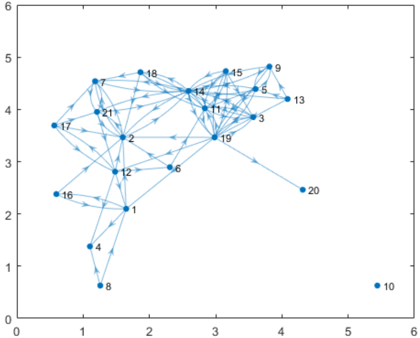
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



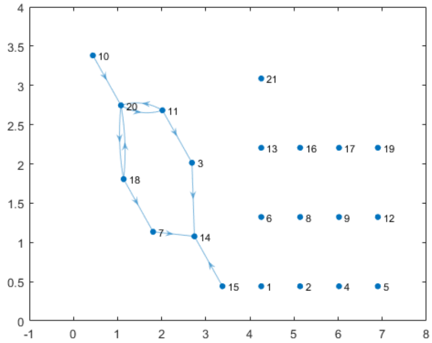
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



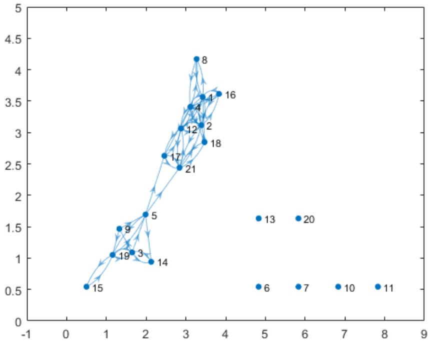
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



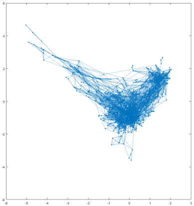
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”

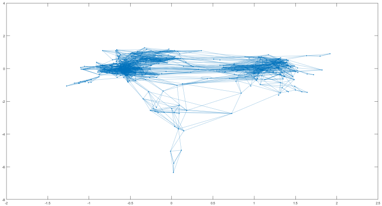


Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Karnataka Village Data



Village 5



Village 61

Figure: Two of the Karnataka Village networks[1]

[1] A. Banerjee, A.G. Chandrasekhar, E. Duflo, and M.O. Jackson, The Diffusion of Microfinance. Science, (2013).

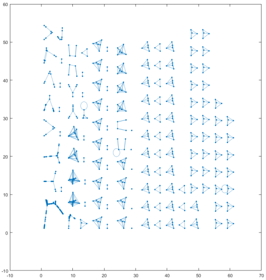
Village Layers

Layer Description	Village 4			Village 61		
	Density	Comp.	Giant %	Density	Comp.	Giant %
Borrow Money	.0082	26	.8354	.0108	15	.9188
Give Advice	.0077	49	.5892	.0098	34	.7377
Help Make Decisions	.0076	61	.1277	.0100	24	.8562
Borrow Kerosene or Rice	.0085	21	.8338	.0113	14	.9171
Lend Kerosene or Rice	.0086	22	.8308	.0113	14	.9255
Lend Money	.0081	14	.7908	.0107	17	.9036
Medical Advice	.0075	84	.2938	.0106	14	.9306
Friends	.0089	15	.9277	.0105	22	.8714
Relatives	.0085	29	.7231	.0105	26	.5448
Attend Temple With	.0073	117	.0462	.0089	108	.0372
Visit Their Home	.0087	15	.9185	.0116	11	.9475
Visit Your Home	.0088	16	.9108	.0117	11	.9492
Aggregate	.0121	3	.9862	.0155	8	.9679

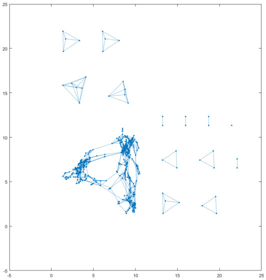
Table: Layer information for two of the Karnataka Villages [1].

[1] D. DeFord and S. Pauls, A new framework for dynamical models on multiplex networks, Journal of Complex Networks, 6(3), 353–381, 2018.

Medical Advice

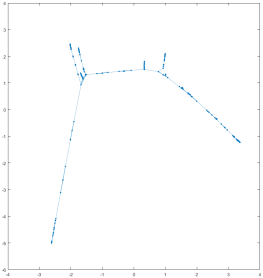


Village 5

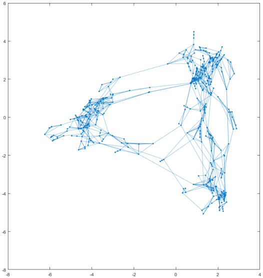


Village 61

Medical Advice

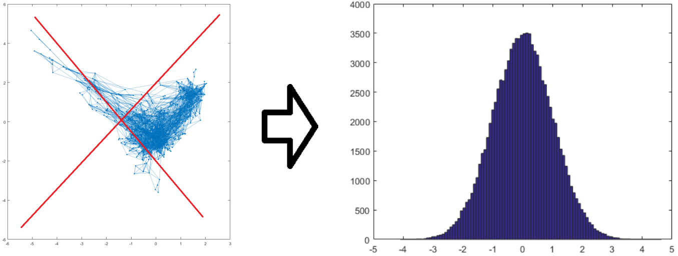


Village 5



Village 61

Random Models



Null Models

Definition (Null Model)

A random network, parameterized to match some features of a given network, used to compare "expected" network measures.

Null Models

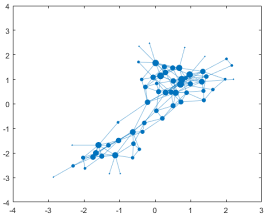
Definition (Null Model)

A random network, parameterized to match some features of a given network, used to compare “expected” network measures.

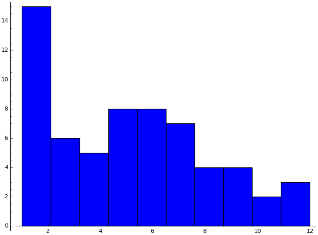
Example (How to grow a network?)

- Ego networks
  - Start with a single node
  - Attach that node to k other nodes
  - Add edges between the each pair of friends with probability p
- ?

Dolphins Degree Distribution



Dolphins



Degree Distribution

Erdos–Renyi (Independence)

Parameters

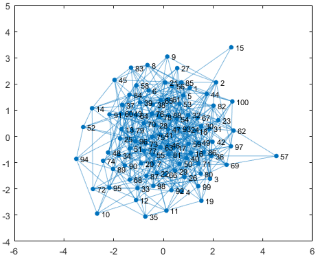
- Number of nodes n
- Connection Probability: p

Process

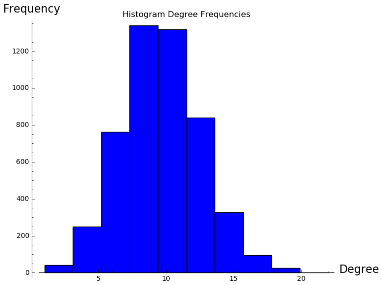
- Start with n nodes
- Connect each pair of nodes independently with probability p

P. Erdos and A. Renyi: On Random Graphs. I, Publicationes Mathematicae, 6, 290-–297, (1959).

Erdos–Renyi (Independence)

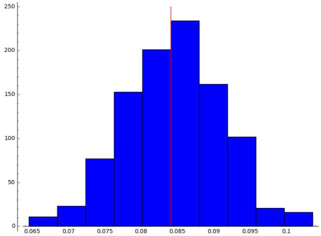


Graph Example

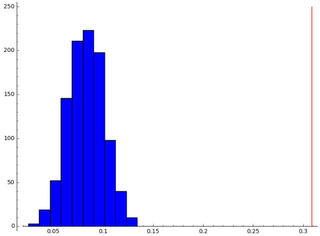


Degree Distribution

ER Transitivity Comparison



Density



Transitivity

Dynamics on Networks

- **Types of Dynamics**
  - Information flow
  - Opinion/consensus diffusion
  - Epidemiology
- **Dynamic Metrics**
  - Centrality
  - Clustering
  - Evolution
- **Animations!**

Computation Time!

- Head to github.com/drdeford/CISER2023NetworkX
- We'll start by making sure that everyone has Python installed or is ready to run notebooks on the cloud somewhere
- Then we'll work through the notebooks - lots of time for questions/discussion

Configuration Model (Degree Sequence)

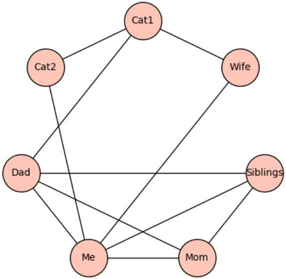
Parameters

- Initial graph

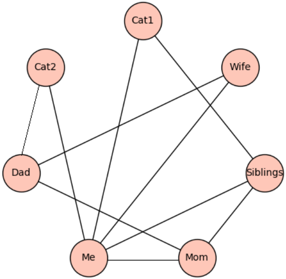
Process

- Start with the initial graph
- Cut each edge in half so that each end remains attached to one of the adjacent nodes
- Randomly pair up the half–edges

Configuration Model (Degree Sequence)

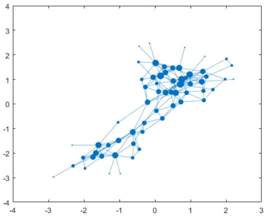


Graph Example

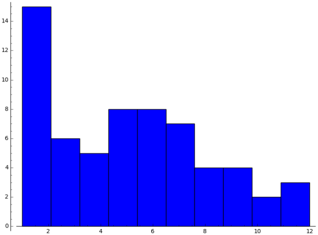


Rewired

Dolphins Degree Distribution

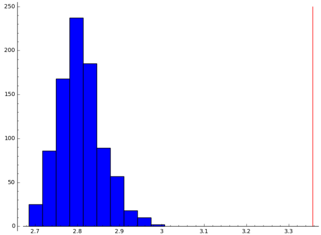


Dolphins

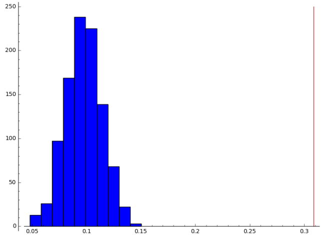


Degree Distribution

Configuration Transitivity Comparison



Average Path Length



Transitivity

Erdos–Renyi (Independence)

Parameters

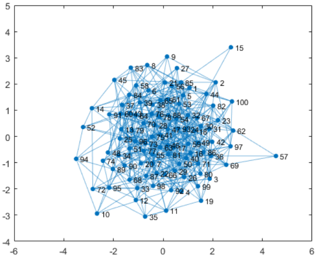
- Number of nodes n
- Connection Probability: p

Process

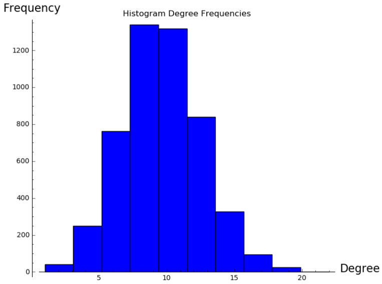
- Start with n nodes
- Connect each pair of nodes independently with probability p

P. Erdos and A. Renyi: On Random Graphs. I, Publicationes Mathematicae, 6, 290-–297, (1959).

Erdos–Renyi (Independence)

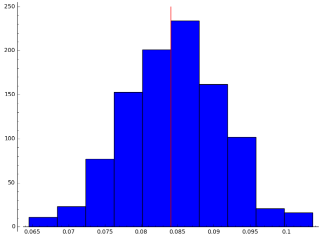


Graph Example

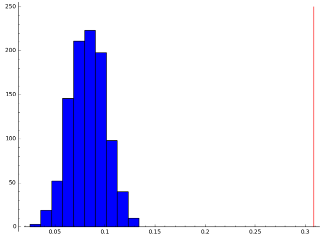


Degree Distribution

ER Transitivity Comparison



Density



Transitivity

Barabasi–Albert (Centrality)

Parameters

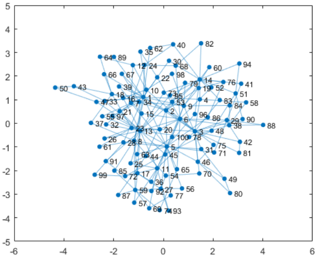
- Initial graph
- Number of nodes: n
- Number of neighbors for new nodes: m

Process

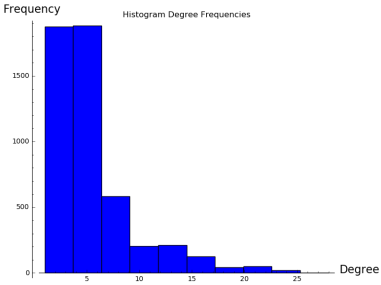
- Start with the initial graph
- Add nodes one at a time until there are n total
- Each added node gets connected to m nodes already in the graph
- These connections are chosen so that the probability that the new node is connected to an existing node is proportional to the degree of the existing node

A. Barabasi and R. Albert: Emergence of scaling in random networks, Science, 286 (5439), 509-–512, (1999).

Barabasi–Albert (Centrality)

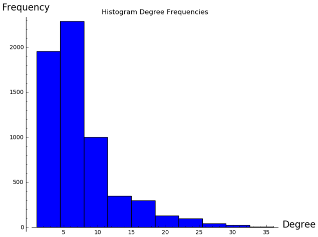


Graph Example

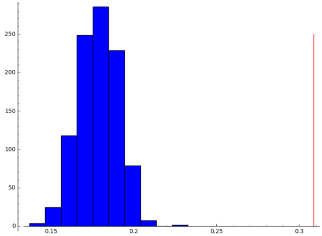


Degree Distribution

BA Transitivity Comparison



Degrees



Transitivity

Watts–Strogatz (Local Clustering)

Parameters

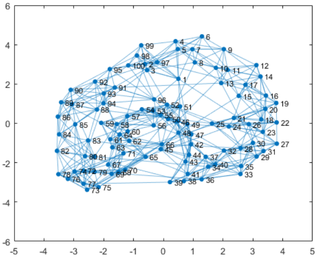
- Number of nodes n
- Number of initial neighbors: k
- Rewiring Probability: p

Process

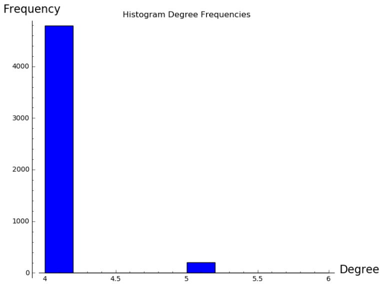
- Start with n nodes connected in a ring so that each node is connected to $\frac{k}{2}$ nodes on each side
- For each edge in the initial graph, rewire it with probability p to a uniformly chosen other node in the graph

D. Watts and S. Strogatz, Collective dynamics of 'small-world' networks, Nature, 393 (6684), 440-–442, (1998).

Watts–Strogatz (Local Clustering)

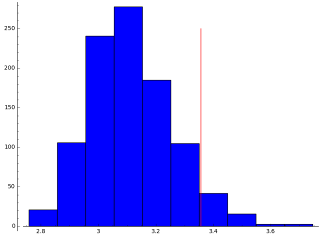


Graph Example

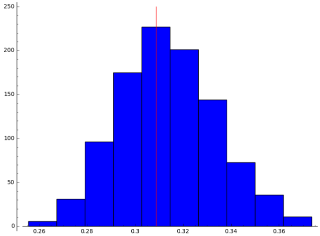


Degree Distribution

WS Transitivity Comparison



Average Path Length



Transitivity

Ego Models

Example (Things to think about)

- https://math.wsu.edu/ddeford/STS_Tufts
- Can you generate your ego network with one of these null models?
- Which model is most likely to generate your ego network?
- Which features of your network aren't described well by any of the models?
- Can you create your own model that generates networks similar to your ego network?

Summary so far

Random network models are designed to replicate (or provide generative stories for) features commonly observed in social networks:

- Short average path length (Erdos–Renyi)
- High Clustering Coefficient (Watts-Strogattz)
- Power law degree distribution (Barabasi-Albert)
- Exact (expected) match of degree distribution (Cofiguration Model or Chung-Lu)

However, none of these models incorporate another important aspect of social networks - community structure.

SBM Basics

- Parameters: n, k, z, M

SBM Basics

- Parameters: n, k, z, M
- k is a fixed number of communities and z is a $1 \times n$ vector recording the assignment of each node to a single community

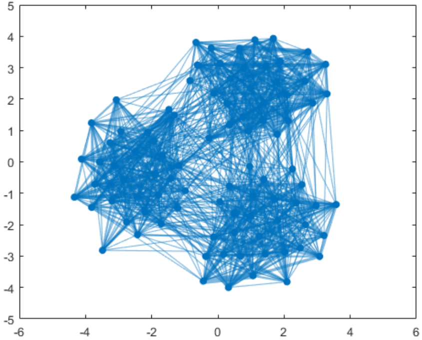
SBM Basics

- Parameters: n, k, z, M
- k is a fixed number of communities and z is a $1 \times n$ vector recording the assignment of each node to a single community
- The entries of $M_{i,j}$ are the (Bernoulli) probabilities that an arbitrary node in community i has an edge to an arbitrary node in j

SBM Basics

- Parameters: n, k, z, M
- k is a fixed number of communities and z is a $1 \times n$ vector recording the assignment of each node to a single community
- The entries of $M_{i,j}$ are the (Bernoulli) probabilities that an arbitrary node in community i has an edge to an arbitrary node in j
- To build the network we flip $\binom{n}{2}$ weighted coins with probabilities given by M_{z_ℓ, z_k}

Assortative SBM Example



MLE Digression

- For the Erdos-Renyi model, each edge is independent with some probability p, so the number of edges in the graph can be viewed as a binomial random variable with $\binom{n}{2}$ trials and the proportion of observed edges is the MLE.

MLE Digression

- For the Erdos-Renyi model, each edge is independent with some probability p, so the number of edges in the graph can be viewed as a binomial random variable with $\binom{n}{2}$ trials and the proportion of observed edges is the MLE.
- Once the number of clusters and the node assignments have been fixed, each entry of the matrix M can be estimated the same way $(\hat{M}_{i,j} = \frac{E_{i,j}}{N_i \cdot N_j})$

MLE Digression

- For the Erdos-Renyi model, each edge is independent with some probability p, so the number of edges in the graph can be viewed as a binomial random variable with $\binom{n}{2}$ trials and the proportion of observed edges is the MLE.
- Once the number of clusters and the node assignments have been fixed, each entry of the matrix M can be estimated the same way $(\hat{M}_{i,j} = \frac{E_{i,j}}{N_i \cdot N_j})$
- This leads us to a log-likelihood expression:

$$\sum_{i,j} E_{i,j} \ln(E_{i,j} + (N_i \cdot N_j - E_{i,j}) \ln(N_i \cdot N_j - E_{i,j} - N_i \cdot N_j \ln(N_i \cdot N_j)$$

MLE Digression

- For the Erdos-Renyi model, each edge is independent with some probability p, so the number of edges in the graph can be viewed as a binomial random variable with $\binom{n}{2}$ trials and the proportion of observed edges is the MLE.
- Once the number of clusters and the node assignments have been fixed, each entry of the matrix M can be estimated the same way ($\hat{M_{i,j}} = \frac{E_{i,j}}{N_i \cdot N_j}$)
- This leads us to a log-likelihood expression:

$$\sum_{i,j} E_{i,j} \ln(E_{i,j} + (N_i \cdot N_j - E_{i,j}) \ln(N_i \cdot N_j - E_{i,j} - N_i \cdot N_j \ln(N_i \cdot N_j)$$

- Gory algebraic details here: `http://tuvalu.santafe.edu/~aaronc/courses/5352/fall2013/csci5352_2013_L17.pdf`

MLE Digression

- For the Erdos-Renyi model, each edge is independent with some probability p, so the number of edges in the graph can be viewed as a binomial random variable with $\binom{n}{2}$ trials and the proportion of observed edges is the MLE.
- Once the number of clusters and the node assignments have been fixed, each entry of the matrix M can be estimated the same way $(\hat{M_{i,j}} = \frac{E_{i,j}}{N_i \cdot N_j})$
- This leads us to a log-likelihood expression:

$$\sum_{i,j} E_{i,j} \ln(E_{i,j} + (N_i \cdot N_j - E_{i,j}) \ln(N_i \cdot N_j - E_{i,j} - N_i \cdot N_j \ln(N_i \cdot N_j)$$

- Gory algebraic details here: `http://tuvalu.santafe.edu/~aaronc/courses/5352/fall2013/csci5352_2013_L17.pdf`
- What about non-Binomial degree distributions: `https://arxiv.org/abs/1008.3926`

Dot Product Graphs

Definition (Dot Product Graph)

G is a dot product graph of dimension d if there exists a map $f : V(G) \to \mathbb{R}^d$ such that $(i, j) \in E(G)$ if and only if $\langle f(i), f(j) \rangle > 1$.

---

[1] C. Fiduccia, E. Scheinerman, A. Trenk, and J. Zito: Dot Product Representations of Graphs, Discrete Mathematics, 181, 1998, 113–138.

[2] R. Kang, L. Lovasz, T. Muller, and E. Scheinerman: Dot Product Representations of Planar Graphs, Electronic Journal of Combinatorics, 18, (2011), 1–14.

[3] Sphere and Dot Product Representations of Graphs: Discrete Computational Geometry, 47, (2012), 548–568.

[4] B. Li and G. Chang: Dot Product Dimension of Graphs, Discrete Applied Mathematics, 166, (2014), 159–163

Dot Product Graphs

Definition (Dot Product Graph)

G is a dot product graph of dimension d if there exists a map $f : V(G) \to \mathbb{R}^d$ such that $(i, j) \in E(G)$ if and only if $\langle f(i), f(j)\rangle > 1$.

- Initial work: Fiduccia et al. (1998)[1]
- Planar graphs: Kang et al. (2011)[2]
- NP–Hard: Kang and Muller (2012)[3]
- $\frac{n}{2}$ critical graphs: Li and Chang (2014)[4]

[1] C. Fiduccia, E. Scheinerman, A. Trenk, and J. Zito: Dot Product Representations of Graphs, Discrete Mathematics, 181, 1998, 113–138.

[2] R. Kang, L. Lovasz, T. Muller, and E. Scheinerman: Dot Product Representations of Planar Graphs, Electronic Journal of Combinatorics, 18, (2011), 1–14.

[3] Sphere and Dot Product Representations of Graphs: Discrete Computational Geometry, 47, (2012), 548–568.

[4] B. Li and G. Chang: Dot Product Dimension of Graphs, Discrete Applied Mathematics, 166, (2014), 159–163

RDPM

(RDPM 1): Select the number of nodes n.

RDPM

(RDPM 1): Select the number of nodes n.

(RDPM 2): Select the latent dimension d.

RDPM

(RDPM 1): Select the number of nodes n.

(RDPM 2): Select the latent dimension d.

(RDPM 3): Select a distribution W over $\mathbb{R}^d$ with $\mathbb{P}\left(\langle X, Y\rangle \in (0,1)\right) = 1$ where X and Y are drawn independently from W.

RDPM

(RDPM 1): Select the number of nodes n.

(RDPM 2): Select the latent dimension d.

(RDPM 3): Select a distribution W over $\mathbb{R}^d$ with $\mathbb{P}\left(\langle X, Y\rangle \in (0,1)\right) = 1$ where X and Y are drawn independently from W.

(RDPM 4): For each node, $1 \leq j \leq n$, draw a vector, $X_j \in \mathbb{R}^d$ from W.

RDPM

(RDPM 1): Select the number of nodes n.

(RDPM 2): Select the latent dimension d.

(RDPM 3): Select a distribution W over $\mathbb{R}^d$ with $\mathbb{P}\left(\langle X, Y\rangle \in (0,1)\right) = 1$ where X and Y are drawn independently from W.

(RDPM 4): For each node, $1 \leq j \leq n$, draw a vector, $X_j \in \mathbb{R}^d$ from W.

(RDPM 5): Form an adjacency matrix, A, form a network with $A_{j,\ell}$ drawn from $\mathrm{Bernoulli}(\langle X_j, X_\ell\rangle)$ for $j \neq \ell$ and $A_{j,j} = 0$ for all $1 \leq j \leq n$.

Interpretations

- Since each node is associated to a vector, it is natural to try and interpret the properties of the node from the vector

Interpretations

- Since each node is associated to a vector, it is natural to try and interpret the properties of the node from the vector
- $\langle x, y \rangle = ||x|| \cdot ||y|| \cos(x, y)$

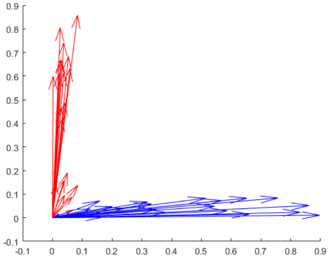
Interpretations

- Since each node is associated to a vector, it is natural to try and interpret the properties of the node from the vector
- $\langle x, y \rangle = ||x|| \cdot ||y|| \cos(x, y)$
- Angle – Community assignment

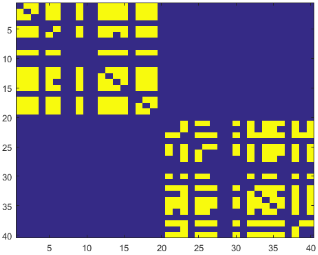
Interpretations

- Since each node is associated to a vector, it is natural to try and interpret the properties of the node from the vector
- $\langle x, y \rangle = ||x|| \cdot ||y|| \cos(x, y)$
- Angle – Community assignment
- Magnitude – Centrality

Angle – Community Assignment

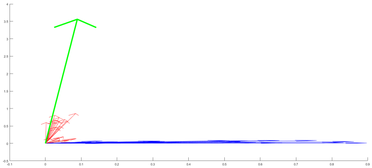


Vectors

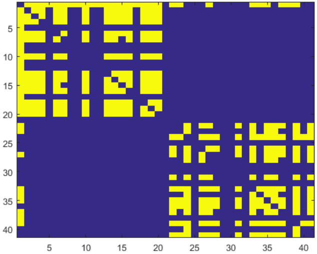


Graph

Magnitude – Centrality



Vectors



Graph

Network Properties

- Initial work: Kraetzel et al. (2005)[5]
- General distributions: Young and Sceinerman (2007)[6]

[5] M. Kraetzel, C. Nickel, and E. Scheinerman: *Random Dot Product Networks: A model for social networks*, Preliminary Manuscript, (2005).

[6] S. Young and E. Scheinerman: *Random Dot Product Models for Social Networks*, Algorithms and Models for the Web-Graph, Lecture Notes in Computer Science, 4863, (2007), 138–149.

Network Properties

- Initial work: Kraetzel et al. (2005)[5]
- General distributions: Young and Sceinerman (2007)[6]
- Small world networks
  - Clustering
  - Small diameter
  - Degree distribution

[5] M. Kraetzel, C. Nickel, and E. Scheinerman: *Random Dot Product Networks: A model for social networks*, Preliminary Manuscript, (2005).

[6] S. Young and E. Scheinerman: *Random Dot Product Models for Social Networks*, Algorithms and Models for the Web-Graph, Lecture Notes in Computer Science, 4863, (2007), 138–149.

Statistical Applications

- Inverse problem: Scheinerman and Tucker (2010)[7]
  - Iterative SVD for approximating $A_{i,j} = \langle X_i, X_j \rangle$
  - Angular k-means

[7] E. Scheinerman and K. Tucker: *Modeling graphs using dot product representations*, Computational Statistics, 25, (2010), 1–16.

[8] D. Sussman, M. Tang, D. Fishkind, and C. Priebe: *A consistent adjacency spectral embedding for stochastic blockmodel graphs*, Journal of the American Statistical Association, 107, (2012), 1119–1128.

[9] M. Tang, A. Athreya, D. Sussman, V. Lyzinski, and C. Priebe: *A nonparametric two–sample hypothesis testing problem for random graphs*, Arxiv: 1409.2344v2, (2014), 1–24.

[10] M. Tang and C. Priebe: Limit theorems for eigenvectors of the normalized Laplacian for random graphs, ArXiv:1607.08601, (2016), 1–52.

Statistical Applications

- Inverse problem: Scheinerman and Tucker (2010)[7]
  - Iterative SVD for approximating $A_{i,j} = \langle X_i, X_j \rangle$
  - Angular k-means
- Spectral Embedding and Statistics: Priebe Lab (2012–present)
  - Adjacency embedding[8]
  - Hypothesis testing[9]
  - Limit theorems[10]

---

[7] E. Scheinerman and K. Tucker: *Modeling graphs using dot product representations*, Computational Statistics, 25, (2010), 1–16.

[8] D. Sussman, M. Tang, D. Fishkind, and C. Priebe: *A consistent adjacency spectral embedding for stochastic blockmodel graphs*, Journal of the American Statistical Association, 107, (2012), 1119–1128.

[9] M. Tang, A. Athreya, D. Sussman, V. Lyzinski, and C. Priebe: *A nonparametric two–sample hypothesis testing problem for random graphs*, Arxiv: 1409.2344v2, (2014), 1–24.

[10] M. Tang and C. Priebe: Limit theorems for eigenvectors of the normalized Laplacian for random graphs, ArXiv:1607.08601, (2016), 1–52.

Advertisement

https://arxiv.org/abs/1611.02530