Mathematical Modeling of Social Connections

Daryl DeFord

MIT/Tufts
Metric Geometry and Gerrymandering Group

STS Lunch Seminar
Tufts University
September 14, 2018

Outline

Code for this talk:

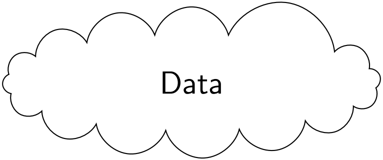
Example (Sage code:)

An interactive tool for experimenting with the objects from this talk can be found here:

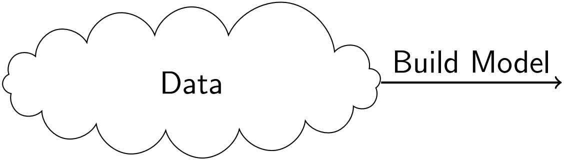
- `https:/people.csail.mit.edu/ddeford/STS_Tufts`
- `https:/people.csail.mit.edu/ddeford/sage_cell_dolphins`

You can right–click to open the page source and copy the Sage code to `https://people.csail.mit.edu.ddeford/sage_cell` to make modifications and run variations.

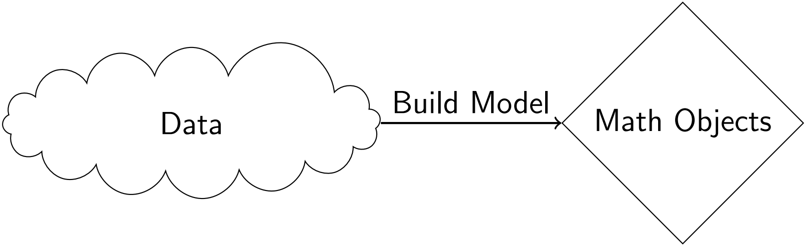
Philosophy



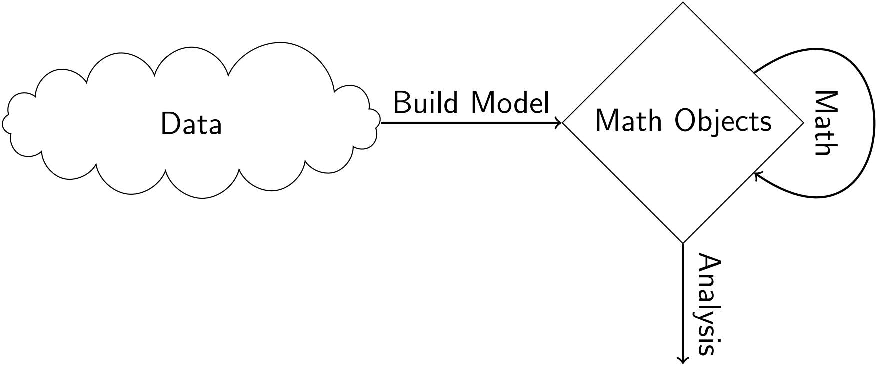
Philosophy



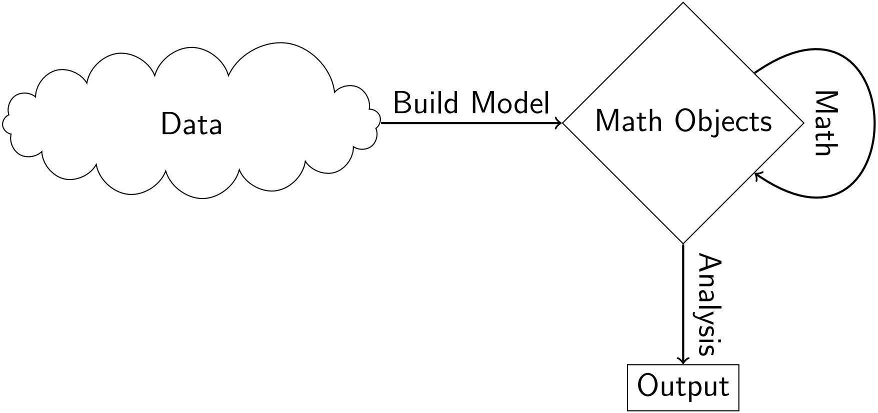
Philosophy



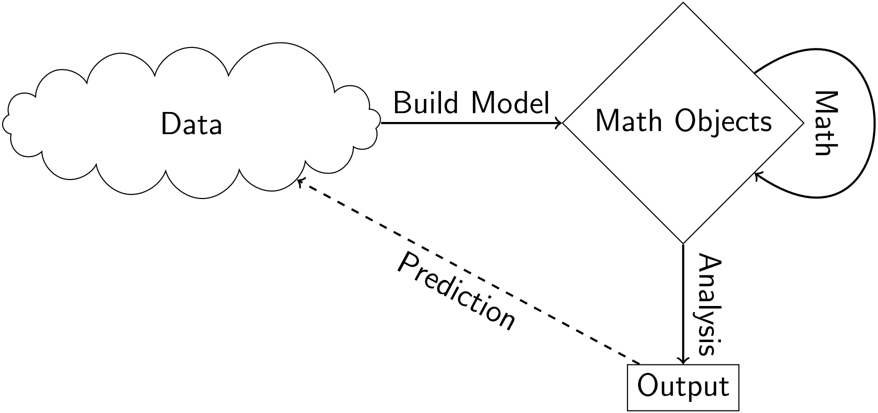
Philosophy



Philosophy



Philosophy


Data
Build Model
Math Objects
Math
Analysis
Output
Prediction

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?
  - **Academic Departments**

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?

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**

What is a social network?

Definition (Social Network)

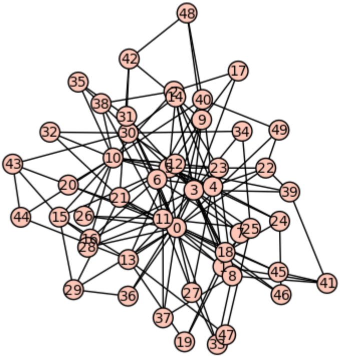
Mathematically, a social network is represented by a collection of "nodes" representing individual actors and a set of "edges" representing a binary relationship between the actors.

Example

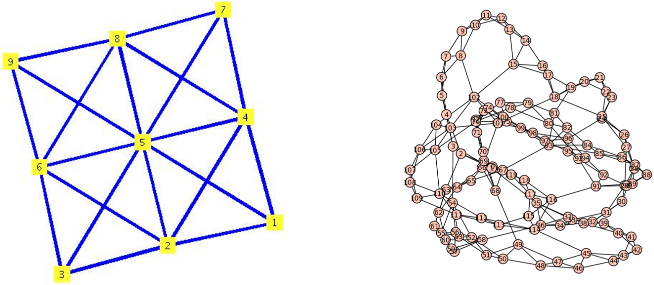
What kinds of systems can social networks describe?

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

Social Networks

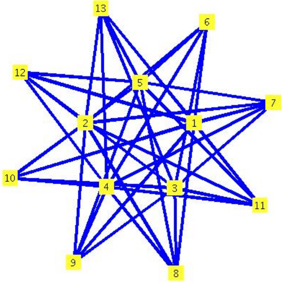


Examples

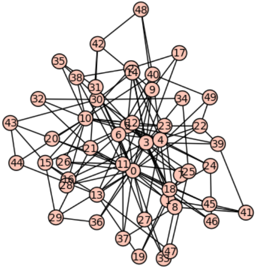


(a) Graph (b) Network

Examples

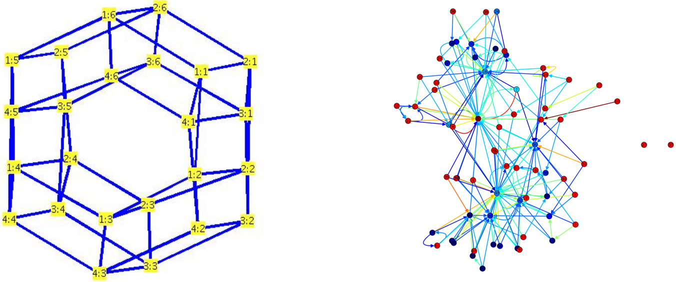


(a) Graph



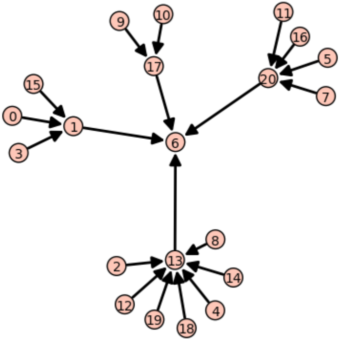
(b) Network

Examples

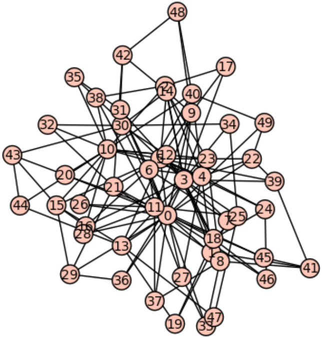


(a) Graph (b) Network

Centrality



Centrality


48
42
17
35
14
40
38
31
9
30
34
49
32
10
43
12
23
22
6
3
4
39
20
21
44
15
26
11
0
24
25
16
28
18
13
45
41
8
46
29
36
27
37
47
19

Centrality

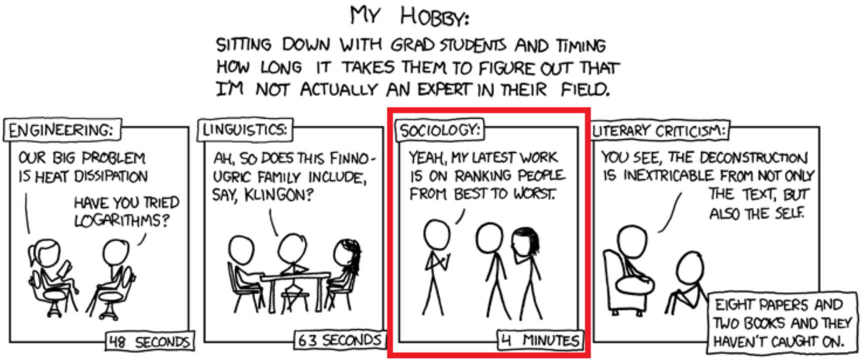
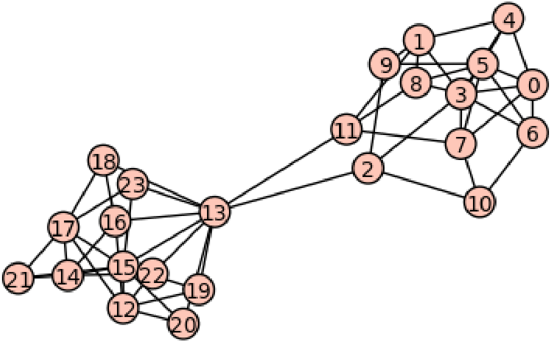


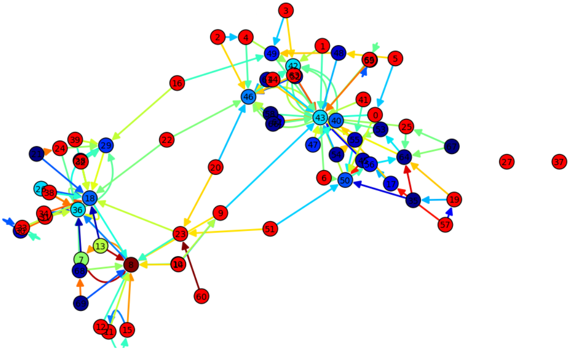
Figure: Relevant comic by Randall Munroe[1] (emphasis mine).

1 https://xkcd.com/451/

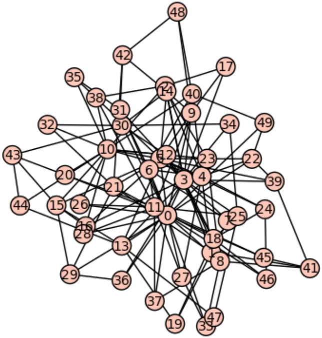
Clustering



Clustering



Clustering


48
42
17
35
14
40
38
31
9
32
30
34
49
43
10
12
23
22
6
3
4
39
20
21
44
15
26
11
0
25
24
16
28
18
13
8
45
41
46
29
36
27
37
47
19

Common Properties of Social Networks

Example (What features distinguish social networks?)

- ?

Common Properties of Social Networks

Example (What features distinguish social networks?)

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

Ego Networks

Definition (Ego Network)

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

Ego Networks

Definition (Ego Network)

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

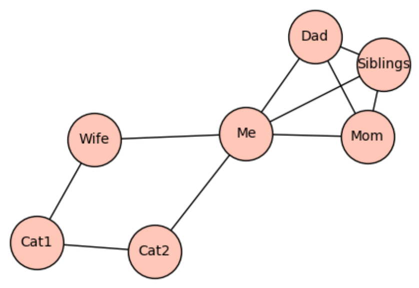
Example (Draw your ego network)

Ego Networks

Definition (Ego Network)

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

Example (Draw your ego network)



How to construct networks?

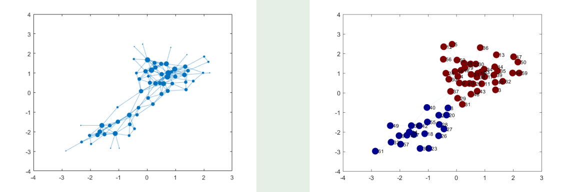
Example (Which edges to add?)

- ?

How to construct networks?

Example (Which edges to add?)

- ?
- Proximity

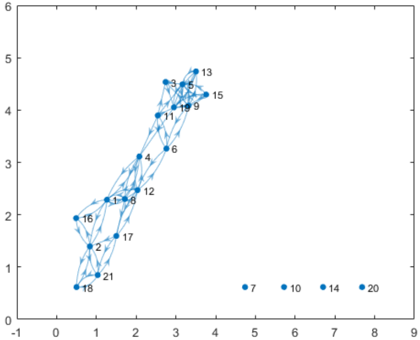


(a) Centrality   (b) Clustering

Figure: Dolphin social network[1]

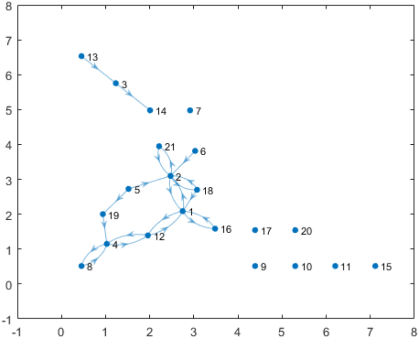
s [1] D. Lusseau, K. Schneider, O. Boisseau, Patti Haase, E. Slooten, and S. Dawson, The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations, Behavioral Ecology and Sociobiology 54 (2003), no. 4, 396–405.

Different Perspectives on “Friendship”



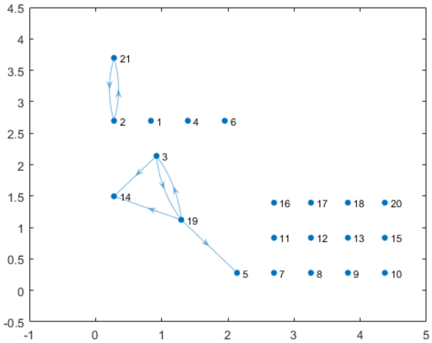
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



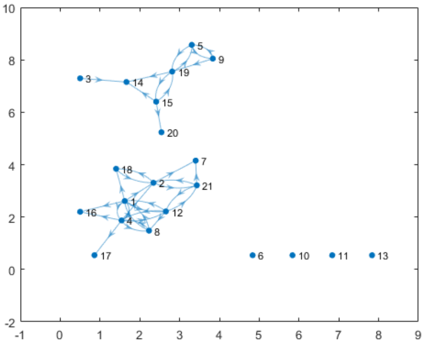
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



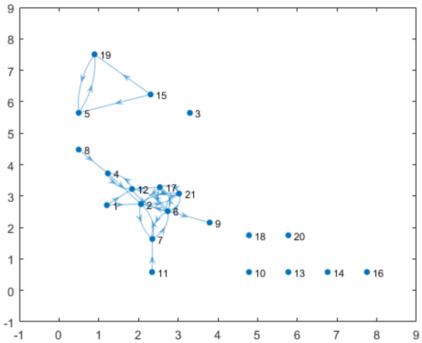
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



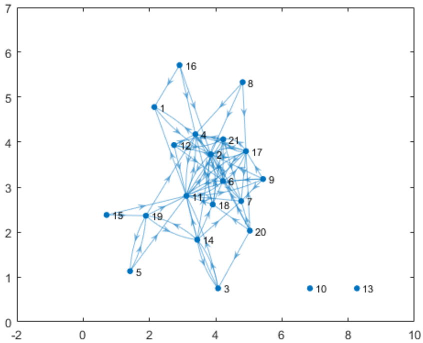
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



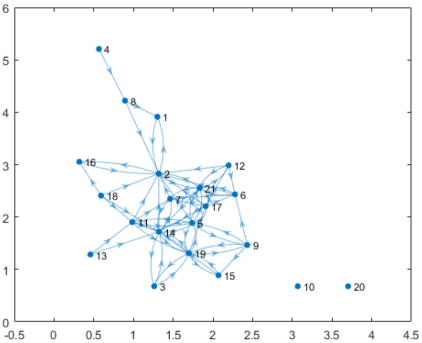
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



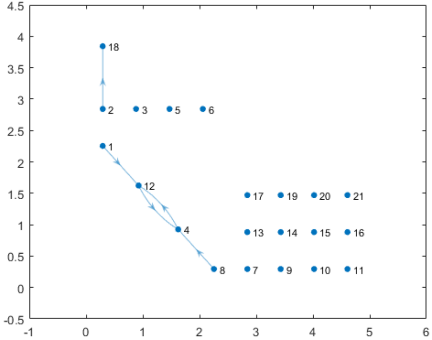
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



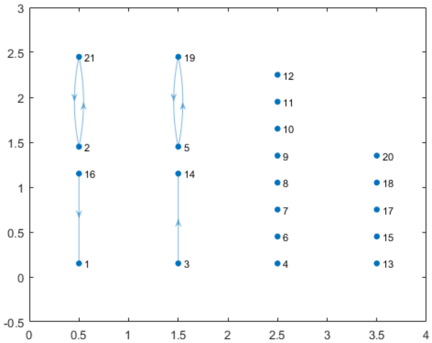
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



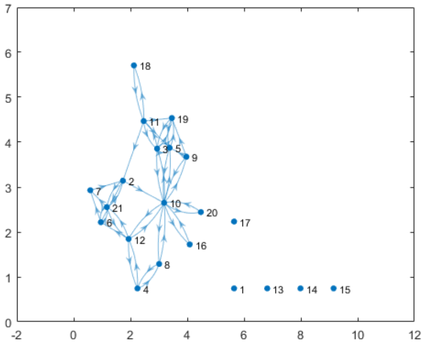
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



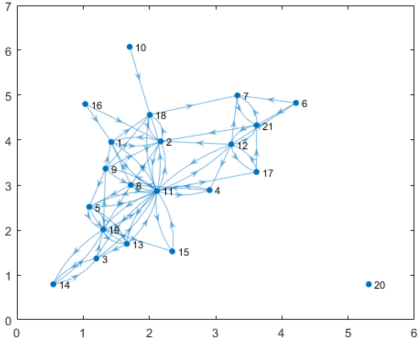
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



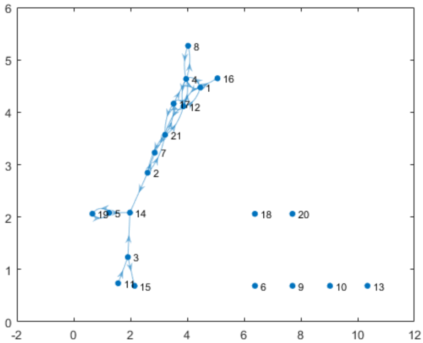
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



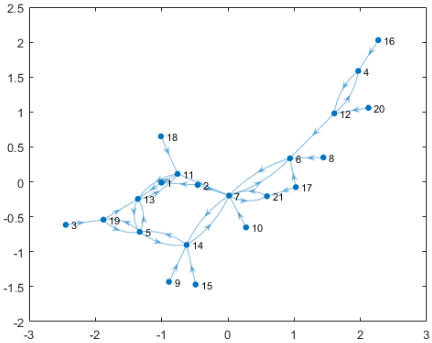
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



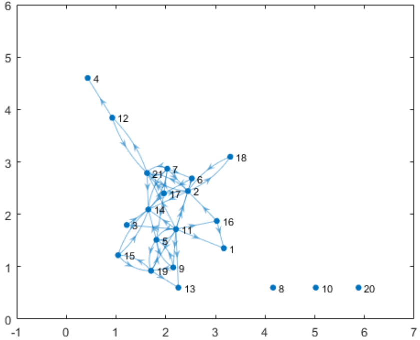
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



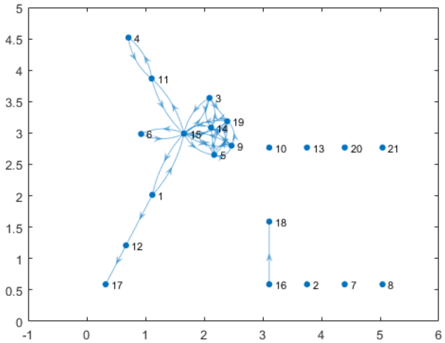
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



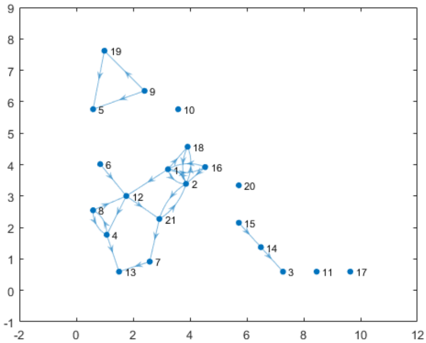
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



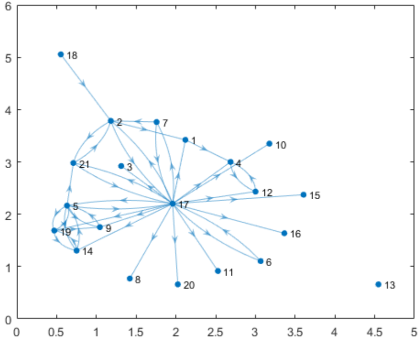
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



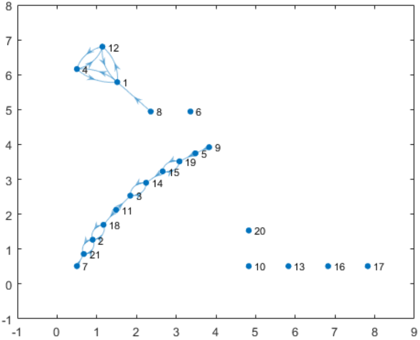
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



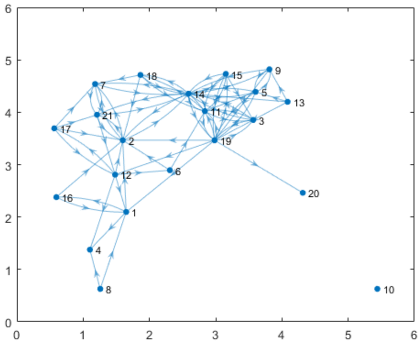
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



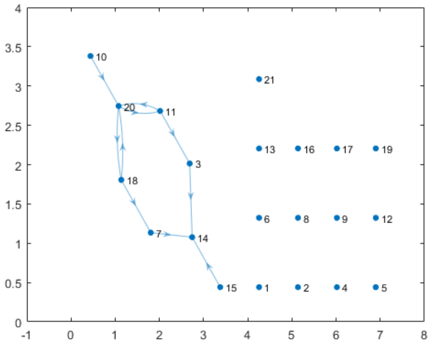
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



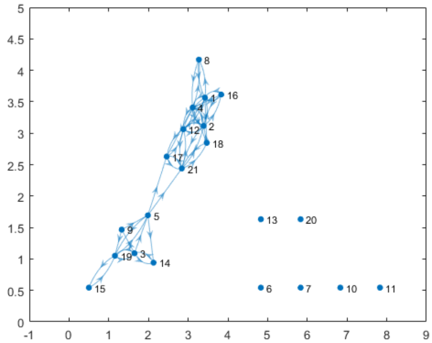
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



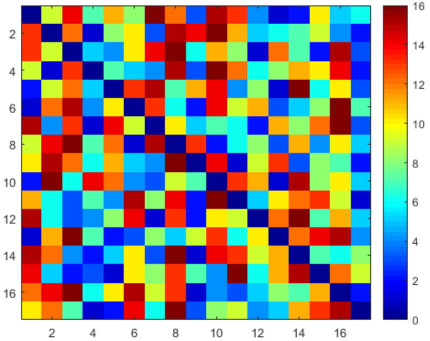
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on “Friendship”



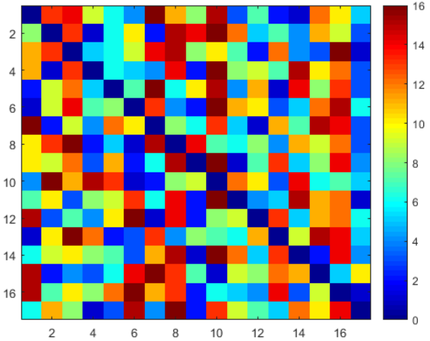
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

"Friendship" over Time



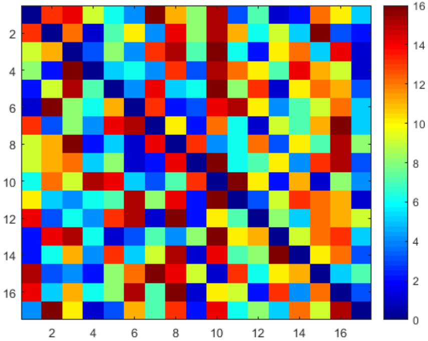
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



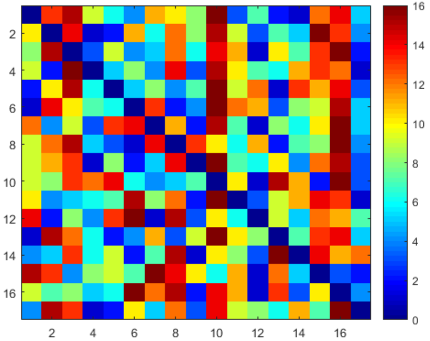
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



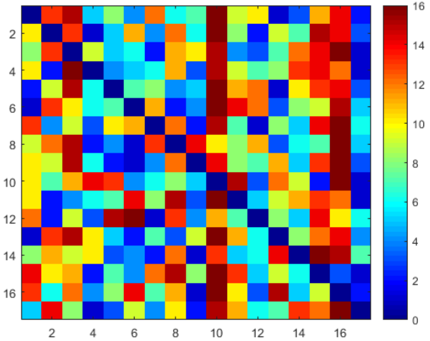
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



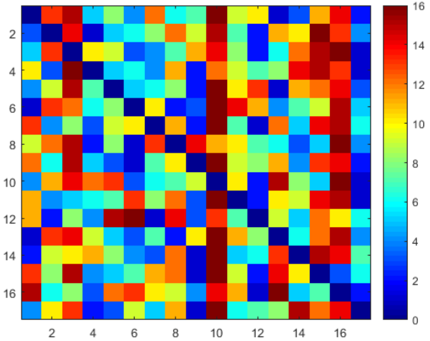
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



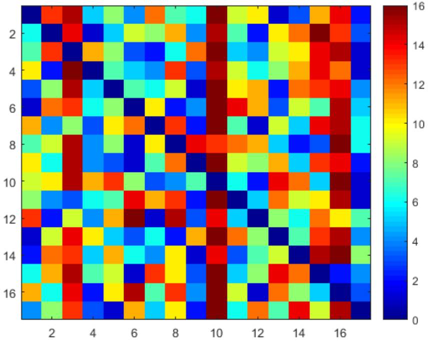
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

“Friendship” over Time



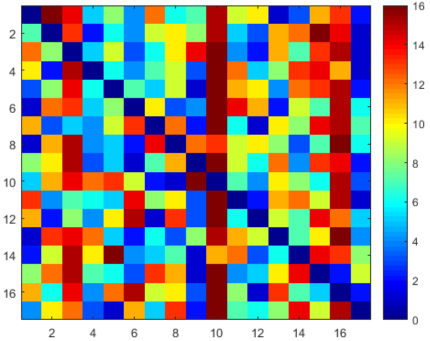
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

“Friendship” over Time



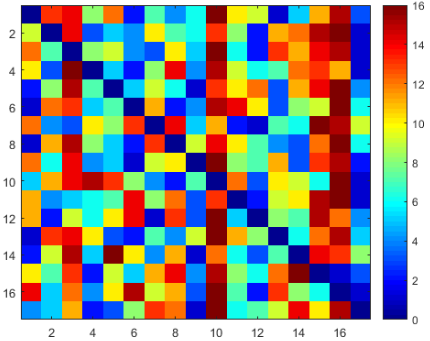
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

“Friendship” over Time



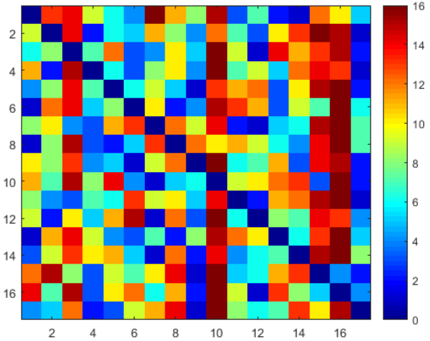
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



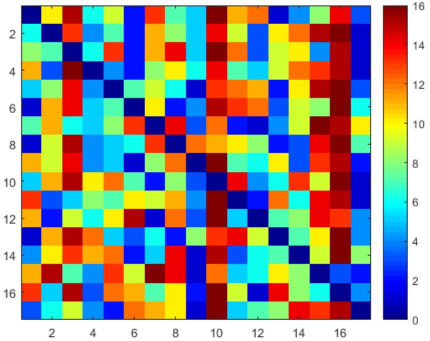
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

“Friendship” over Time



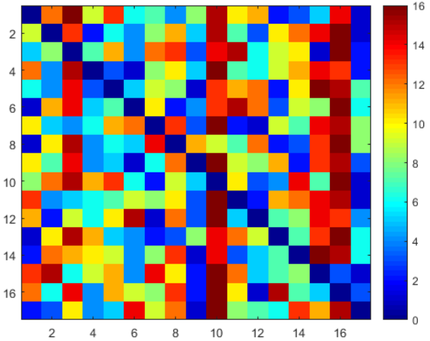
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



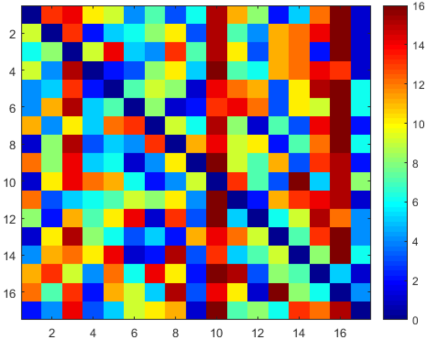
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



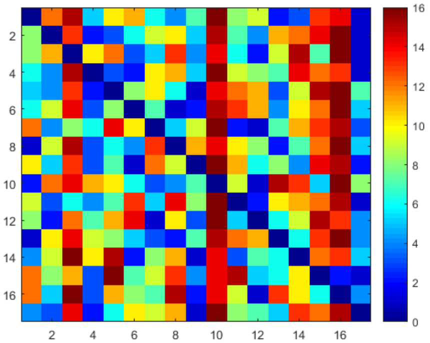
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



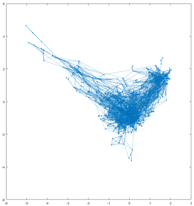
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time

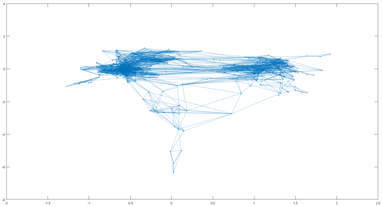


Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

Karnataka Village Data



(a) Village 5



(b) 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).

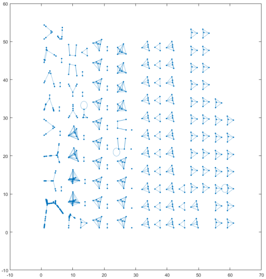
Village Layers

Layer	Village 4			Village 61		
Description	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

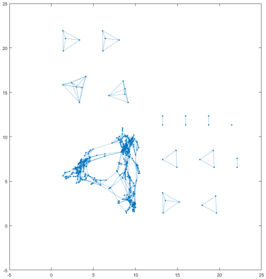
Table: Layer information for two of the Karnataka Villages [1].

[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

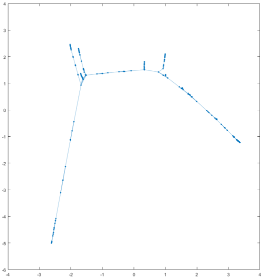


(a) Village 5

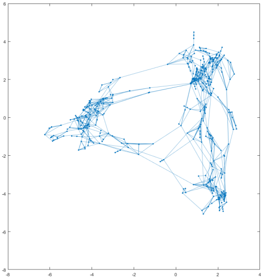


(b) Village 61

Medical Advice

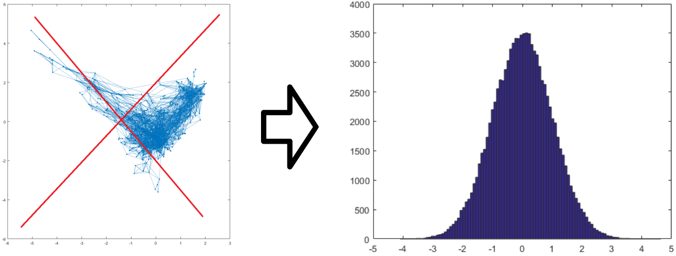


(a) Village 5



(b) 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
- ?

Configuration Model (Degree Sequence)

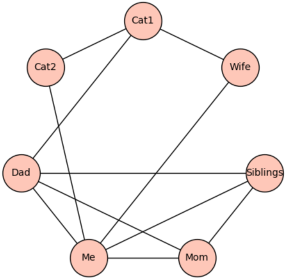
Parameters

- Initial graph

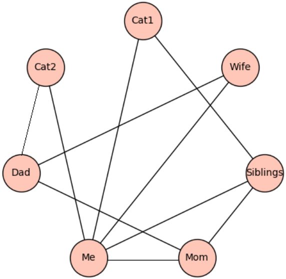
Process

- 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)

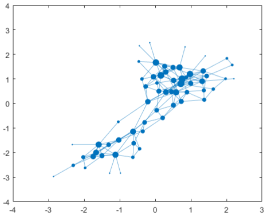


(a) Graph Example

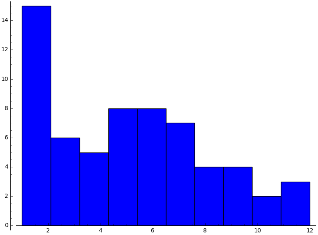


(b) Rewired

Dolphins Degree Distribution

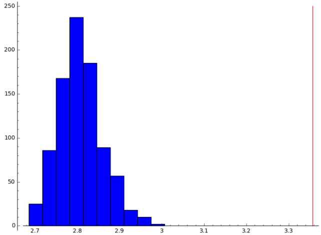


(a) Dolphins

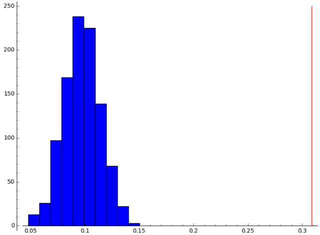


(b) Degree Distribution

Configuration Transitivity Comparison



(a) Average Path Length



(b) Transitivity

Erdos–Renyi (Independence)

Parameters

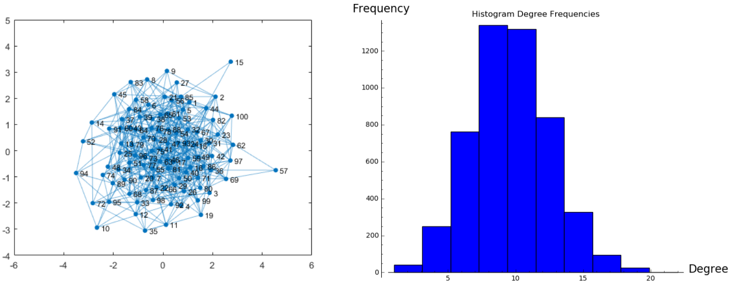
- Number of nodes n
- Connection Probability: p

Process

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

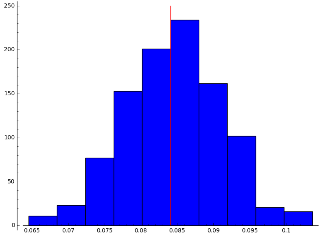
P. Erdos and A. Renyi: On Random Graphs. I, Publicationes Mathematicae, 6, 290-–297, (1959).

Erdos–Renyi (Independence)

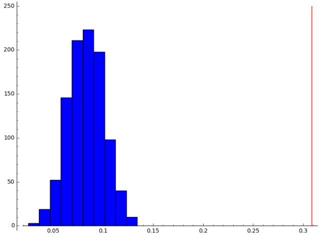


(a) Graph Example (b) Degree Distribution

ER Transitivity Comparison



(a) Density



(b) Transitivity

Barabasi–Albert (Centrality)

Parameters

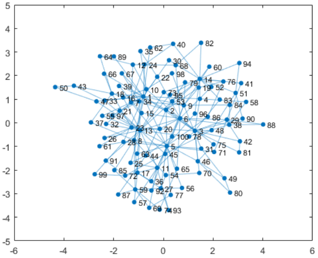
- Initial graph
- Number of nodes: n
- Number of neighbors for new nodes: m

Process

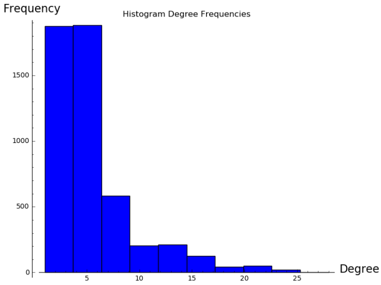
- Start with the initial graph
- Add nodes one at a time until there are n total
- Each added node gets connected to m nodes already in the graph
- These connections are chosen so that the probability that the new node is connected to an existing node is proportional to the degree of the existing node

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

Barabasi–Albert (Centrality)

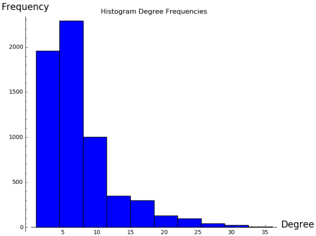


(a) Graph Example

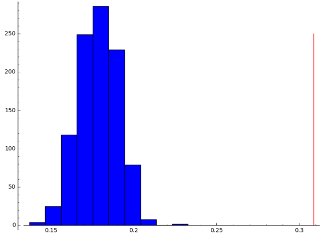


(b) Degree Distribution

BA Transitivity Comparison



(a) Degrees



(b) Transitivity

Watts–Strogatz (Local Clustering)

Parameters

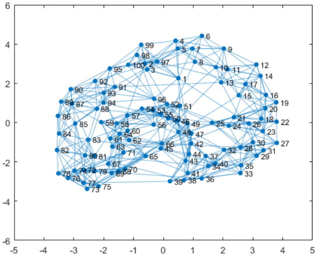
- Number of nodes n
- Number of initial neighbors: k
- Rewiring Probability: p

Process

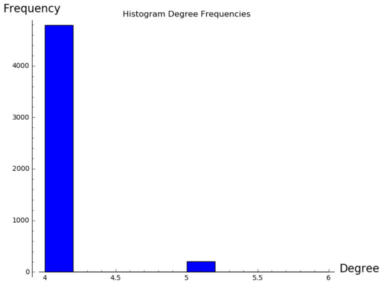
- Start with n nodes connected in a ring so that each node is connected to $\frac{k}{2}$ nodes on each side
- For each edge in the initial graph, rewire it with probability p to a uniformly chosen other node in the graph

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

Watts–Strogatz (Local Clustering)

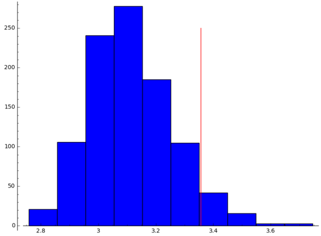


(a) Graph Example

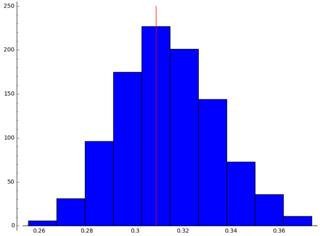


(b) Degree Distribution

WS Transitivity Comparison



(a) Average Path Length



(b) Transitivity

Ego Models

Example (Things to think about)

- `https://people.csail.mit.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?

The end!

Thanks!

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

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
- $\langle x, y \rangle = ||x|| \cdot ||y|| \cos(x, y)$

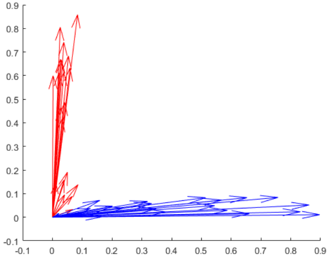
Interpretations

- 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

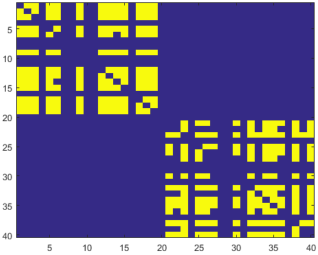
Interpretations

- 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

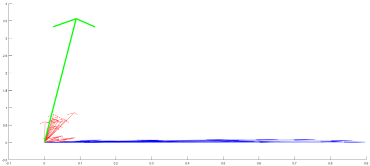


(a) Vectors

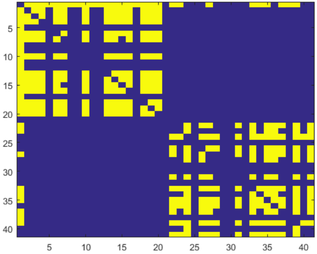


(b) Graph

Magnitude – Centrality



(c) Vectors



(d) Graph

Network Properties

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

WRDPM

(WRDPM 0): Select a probability distribution $P(p_1, p_2, \ldots, p_k)$ for the edge weights. Let $S_i \subseteq \mathbb{R}$ the domain for p_i.

WRDPM

(WRDPM 0): Select a probability distribution $P(p_1, p_2, \ldots, p_k)$ for the edge weights. Let $S_i \subseteq \mathbb{R}$ the domain for p_i.

(WRDPM 1): Select the number of desired nodes n.

WRDPM

(WRDPM 0): Select a probability distribution $P(p_1, p_2, \ldots, p_k)$ for the edge weights. Let $S_i \subseteq \mathbb{R}$ the domain for p_i.

(WRDPM 1): Select the number of desired nodes n.

(WRDPM 2): For each parameter p_i, select a dimension d_i.

(WRDPM 3): For each parameter p_i, select a distribution W_i defined over $\mathbb{R}^{d_i}$ so that $\mathbb{P}(\langle X_i, Y_i \rangle \in S_i) = 1$ where X_i and Y_i are drawn independently from W_i.

WRDPM

(WRDPM 0): Select a probability distribution $P(p_1, p_2, \ldots, p_k)$ for the edge weights. Let $S_i \subseteq \mathbb{R}$ the domain for p_i.

(WRDPM 1): Select the number of desired nodes n.

(WRDPM 2): For each parameter p_i, select a dimension d_i.

(WRDPM 3): For each parameter p_i, select a distribution W_i defined over $\mathbb{R}^{d_i}$ so that $\mathbb{P}(\langle X_i, Y_i \rangle \in S_i) = 1$ where X_i and Y_i are drawn independently from W_i.

(WRDPM 4): For each node, $1 \leq j \leq n$, select k vectors $1 \leq i \leq k$ (one from each parameter space), $X_i^j \in \mathbb{R}^{d_i}$, according to distribution W_i.

WRDPM

(WRDPM 0): Select a probability distribution $P(p_1, p_2, \ldots, p_k)$ for the edge weights. Let $S_i \subseteq \mathbb{R}$ the domain for p_i.

(WRDPM 1): Select the number of desired nodes n.

(WRDPM 2): For each parameter p_i, select a dimension d_i.

(WRDPM 3): For each parameter p_i, select a distribution W_i defined over $\mathbb{R}^{d_i}$ so that $\mathbb{P}(\langle X_i, Y_i \rangle \in S_i) = 1$ where X_i and Y_i are drawn independently from W_i.

(WRDPM 4): For each node, $1 \leq j \leq n$, select k vectors $1 \leq i \leq k$ (one from each parameter space), $X_i^j \in \mathbb{R}^{d_i}$, according to distribution W_i.

(WRDPM 5): Finally, construct a weighted adjacency matrix, A, for the network, with $A_{j,\ell}$ drawn according to $P(\langle X_1^\ell, X_1^j \rangle, \langle X_2^\ell, X_2^j \rangle, \ldots, \langle X_k^\ell, X_k^j \rangle)$ for $j > \ell$, $A_{j,\ell} = A_{\ell,j}$ for $j > \ell$ and $A_{j,j} = 0$ for all $1 \leq j \leq n$.

Example: Uniform Noise

0) Take P to be the Poisson distribution with parameters λ.

Example: Uniform Noise

0) Take P to be the Poisson distribution with parameters λ.
1) Select $n = 150$.

Example: Uniform Noise

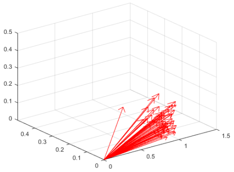
0) Take P to be the Poisson distribution with parameters λ.
1) Select $n = 150$.
2) Choose $d_\lambda = 3$

Example: Uniform Noise

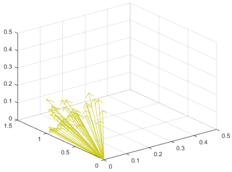
0) Take P to be the Poisson distribution with parameters λ.
1) Select $n = 150$.
2) Choose $d_\lambda = 3$
3) let Y be a normal random variable with mean 0 and variance .1 and take W_λ to be be defined by:

$$W_\lambda = \begin{cases} e_1 + Ye_1 + Ye_2 + Ye_3 & \frac{1}{3} \\ e_2 + Ye_1 + Ye_2 + Ye_3 & \frac{1}{3} \\ e_3 + Ye_1 + Ye_2 + Ye_3 & \frac{1}{3} \end{cases}$$

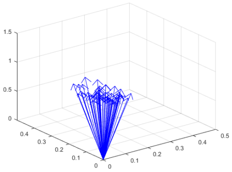
Example:Uniform Noise



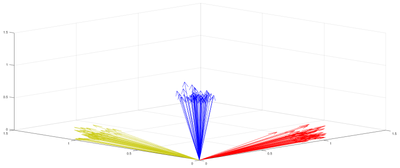
(e) Community 1 Vectors



(f) Community 2 Vectors



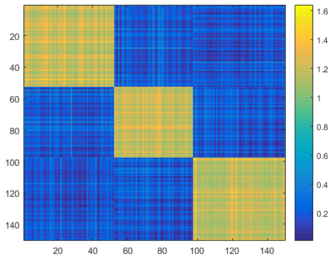
(g) Community 3 Vectors



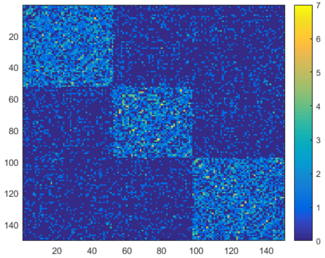
(h) All Vectors

Figure

Example: Uniform Noise



(a) Dot Products



(b) WRDPM Network

Example: Multiresolution Communities

0) Take P to be the Poisson distribution with parameters λ.

Example: Multiresolution Communities

0) Take P to be the Poisson distribution with parameters λ.
1) Select $n = 150$.

Example: Multiresolution Communities

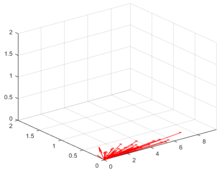
0) Take P to be the Poisson distribution with parameters λ.
1) Select $n = 150$.
2) Choose $d_\lambda = 3$

Example: Multiresolution Communities

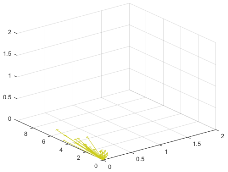
0) Take P to be the Poisson distribution with parameters λ.
1) Select $n = 150$.
2) Choose $d_\lambda = 3$
3) Let X be an exponential random variable with exponent 2, and take W_λ to be be defined by:

$$V_\lambda = \begin{cases} Xe_1 + Ye_2 + Ye_3 & \frac{1}{3} \\ Xe_2 + Ye_1 + Ye_3 & \frac{1}{3} \\ Xe_3 + Ye_1 + Ye_2 & \frac{1}{3} \end{cases}$$

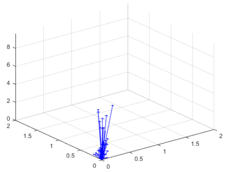
Example: Multiresolution Communities



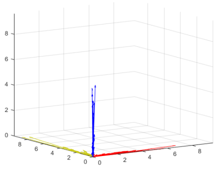
(c) Community 1 Vectors



(d) Community 2 Vectors

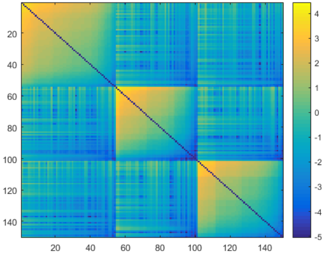


(e) Community 3 Vectors

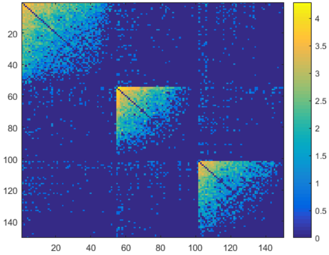


(f) All Vectors

Example: Multiresolution Communities



(g) Dot Products



(h) WRDPM Network

Edge Parameterized Models

Theorem

Let n be a fixed positive integer. For each pair (i, j) with $1 \leq i < j \leq n$ let $a_{i,j} = a_{j,i} \in \mathbb{R}$. Then there exist n real numbers $a_{\ell,\ell}$ for $1 \leq \ell \leq n$ such that the matrix $A_{i,j} = a_{i,j}$ is positive definite.

Proof.

Let the $a_{i,j}$ be selected arbitrarily. For $1 \leq \ell \leq n$ choose $a_{\ell,\ell} \in \mathbb{R}$ so that $a_{\ell,\ell} > \sum_{j \neq \ell} |a_{j,\ell}|$. Form a matrix A with $A_{i,j} = a_{i,j}$. This is a real symmetric matrix and so by the spectral theorem A has real eigenvalues. Applying Gershgorin's Circle Theorem to A gives that the eigenvalues of A lie in the closed disks centered at $a_{\ell,\ell}$ with radius $\sum_{j \neq \ell} |a_{j,\ell}|$. Intersecting these disks with the real line gives that the eigenvalues of A must lie in $\bigcup_{\ell=1}^{n} \left[a_{\ell,\ell} - \sum_{j \neq \ell} |a_{j,\ell}|, a_{\ell,\ell} + \sum_{j \neq \ell} |a_{j,\ell}| \right] \subseteq \mathbb{R}^{+}$. Thus, all eigenvalues of A are positive and A is positive definite. $\square$

Edge Parameterized Models

Corollary

Any generative network model, on a fixed number of nodes n, where the edge weight between each pair of nodes is drawn independently from a fixed probability distribution, possibly with different parameters for each pair, can be realized under the WRDPN.

Proof.

Let P be the k–parameter distribution from which the edge weights are drawn and for $1 \leq i \leq k$ let $a^i_{j,\ell} = a^i_{\ell,j}$ be the value of the ith parameter between nodes j and ℓ. Applying Theorem 1 to the collection $a^i_{j,\ell} = a^i_{\ell,j}$ gives a positive definite matrix A^i. Thus, there exists an $n \times n$ matrix X^i such that $(X^i)^T X^i = A$.
To form the WRDPM that matches the given generative model we take $d_i = n$ for all $1 \leq i \leq k$ and to each node $1 \leq j \leq n$ assign the collection of vectors given by the jth columns of the X^i for $1 \leq i \leq k$. $\square$

Examples

- Erdos–Renyi
  - Single vector for W
  - Simplest null model

[11] J. Ranola, S. Ahn, M. Sehl, D. Smith, and K. Lange: *A Poisson Model for random multigraphs*, Bioinformatics, 26, (2010), 2004–2011.

Examples

- Erdos–Renyi
  - Single vector for W
  - Simplest null model
- Chung–Lu
  - One–dimensional model
  - Expected degree distribution
  - Poisson version: Ranola et al. (2010)[11]

[11] J. Ranola, S. Ahn, M. Sehl, D. Smith, and K. Lange: *A Poisson Model for random multigraphs*, Bioinformatics, 26, (2010), 2004–2011.

Examples

- Erdos–Renyi
  - Single vector for W
  - Simplest null model
- Chung–Lu
  - One–dimensional model
  - Expected degree distribution
  - Poisson version: Ranola et al. (2010)[11]
- WSBM
  - Finite W
  - Community structure
  - Inference

[11] J. Ranola, S. Ahn, M. Sehl, D. Smith, and K. Lange: *A Poisson Model for random multigraphs*, Bioinformatics, 26, (2010), 2004–2011.

Methodology

Want to find a collection of $d \times n$ vectors $\{X_i\}$ in order to approximate the entries of $A_{i,j}$ by $\langle X_i, X_j \rangle$. Equivalently, $X^T X \approx A$.

[12] E. SCHEINERMAN AND K. TUCKER: *Modeling graphs using dot product representations*, Computational Statistics, 25, (2010), 1–16.

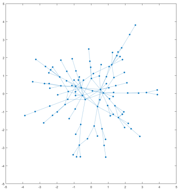
Methodology

Want to find a collection of $d \times n$ vectors $\{X_i\}$ in order to approximate the entries of $A_{i,j}$ by $\langle X_i, X_j \rangle$. Equivalently, $X^T X \approx A$.

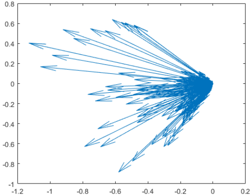
- Positive semi–definite approximation
- Extra degrees of freedom along diagonal
- Introduce a diagonal term
- Alternating, iterative optimization[12]

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

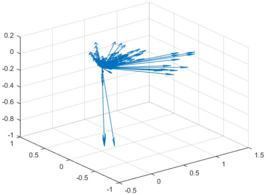
Unweighted Collaboration Network



(i) Collaboration Network



(j) Unweighted 2–Embedding

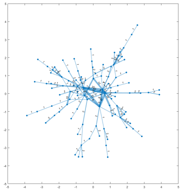


(k) Unweighted 3–Embedding

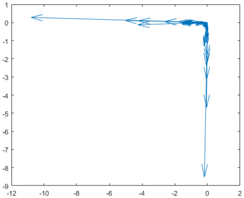
[13]

[13] V. Batagelj and A. Mrvar: *Pajek datasets*, (2006), URL: http://vlado.fmf.uni-lj.si/pub/networks/data/.

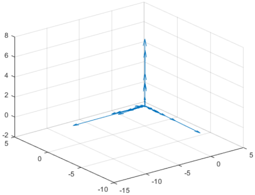
Weighted Collaboration Network



(l) Collaboration Network



(m) Weighted 2–Embedding

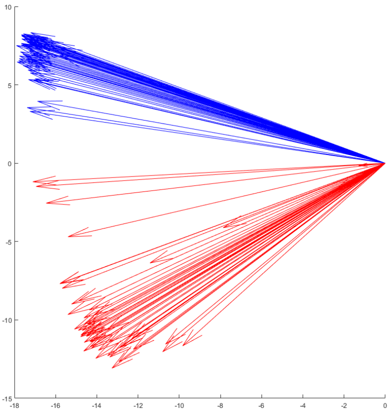


(n) Weighted 3–Embedding

[14]

[14] V. Batagelj and A. Mrvar: *Pajek datasets*, (2006), URL: http://vlado.fmf.uni-lj.si/pub/networks/data/.

Voting Data



J. LEWIS AND K. POOLE: *Roll Call Data*, voteview.com/dwnl.html.

Dimension Selection

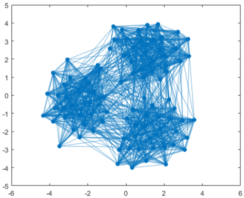
Since the dimension of the embedding is intrinsically related to the realized community structure it is natural to try and make use of this relationship to determine the right choice of d. Motivated by the case of disjoint communities, where if we have an effective, normalized embedding we should have

$$\langle X_i, X_j \rangle = \begin{cases} 1 & \text{i and j belong to the same community} \\ 0 & \text{i and j belong to different communities} \end{cases}$$

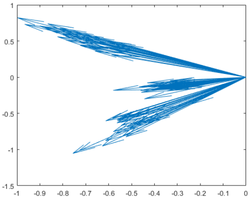
Thus, the sum of intra–community dot products should be $\sum_{i=1}^{\ell} \binom{z_\ell}{2}$. Similarly, the sum of the inter–community dot products should be 0. we define a stress function s depending on the community assignments after embedding.

$$s(d) = \sum_{i=1}^{d} \binom{z_i}{2} - s_{\text{intra}}(d) + s_{\text{inter}}(d)$$

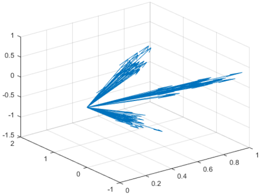
Dimension Example



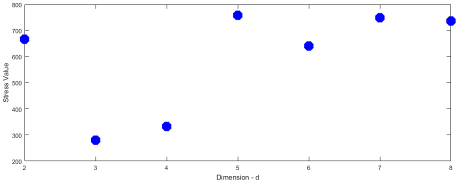
(o) Weighted Network



(p) 2-Embedding



(q) 3-Embedding



(r) Stress Function

Coauthorship Revisited

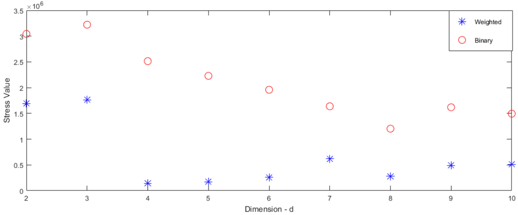


Figure: Comparison of stress values for the computational geometry coauthorship network between the weighted and unweighted realizations. The weighted embedding significantly outperforms the binarized model.