Introduction to Monte Carlo Methods

Daryl DeFord

VRDI
MGGG
June 6, 2018

Outline

Code for this talk:

Remember you can open the page source and copy things to `math.dartmouth.edu/~ddeford/sage_cell` to make modifications and run variations. (links are clickable)

1. `math.dartmouth.edu/~ddeford/war`
2. `math.dartmouth.edu/~ddeford/cube_dist`
3. `math.dartmouth.edu/~ddeford/pisimple`
4. `math.dartmouth.edu/~ddeford/mcmc1`
5. `math.dartmouth.edu/~ddeford/mcmc2`
6. `math.dartmouth.edu/~ddeford/mcmc3`
7. `math.dartmouth.edu/~ddeford/mcmc4`
8. `math.dartmouth.edu/~ddeford/mcmc5`

Warmup Card Problems

1. In a shuffled deck, what is the probability that the top card is red **and** a queen?
2. In a shuffled deck, what is the probability that the top card is red **or** a queen?
3. In a shuffled deck, what is the probability that **at least** one of the top two cards is an ace?
4. In a shuffled deck, what is the probability that **exactly** one of the top two cards is an ace?
5. In a shuffled deck, what is the probability that **at most** one of the top two cards is an ace?
6. In a shuffled deck, what is the probability that a 5 card hand contains at least one card of each suit?
7. In a shuffled deck, what is the probability that a 5 card hand contains no cards of exactly one suit?

Game: Cooperative War

Rules:

1. Nominate a dealer in your group of 3 players
2. Deal 2 cards to each player
3. Each round begins with the dealer playing the highest value card from their hand.
4. The round continues counterclockwise with each player playing the highest card in their hand **only if** it is higher than the previously played card. If not, skip that player and move on to the next.
5. The round ends once each player has had an opportunity to play a card.
6. You (collectively) win if all players have played all of their cards at the end of the second round.

Cooperative War Results?

Question

What is the probability that you win, given a randomly shuffled deck?

Cooperative War Results?

Question

What is the probability that you win, given a randomly shuffled deck?

Answer

Try it out!

Geometric Probability

Question

What is the expected distance between two random points on $[0,1]$*?*

Answer

$$\int_0^1 \int_0^1 |x-y| dx dy = \frac{1}{3}$$

Question

What is the expected distance between two random points on $[0,1]^n$*?*

Answer

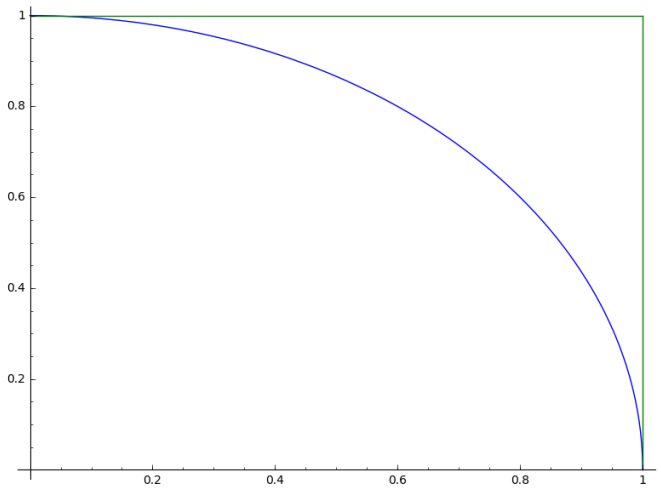
$$\int_0^1 \cdots \int_0^1 \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2} dx_1 \cdots dx_n dy_1 \cdots dy_n = \quad :($$

Numerical Integration

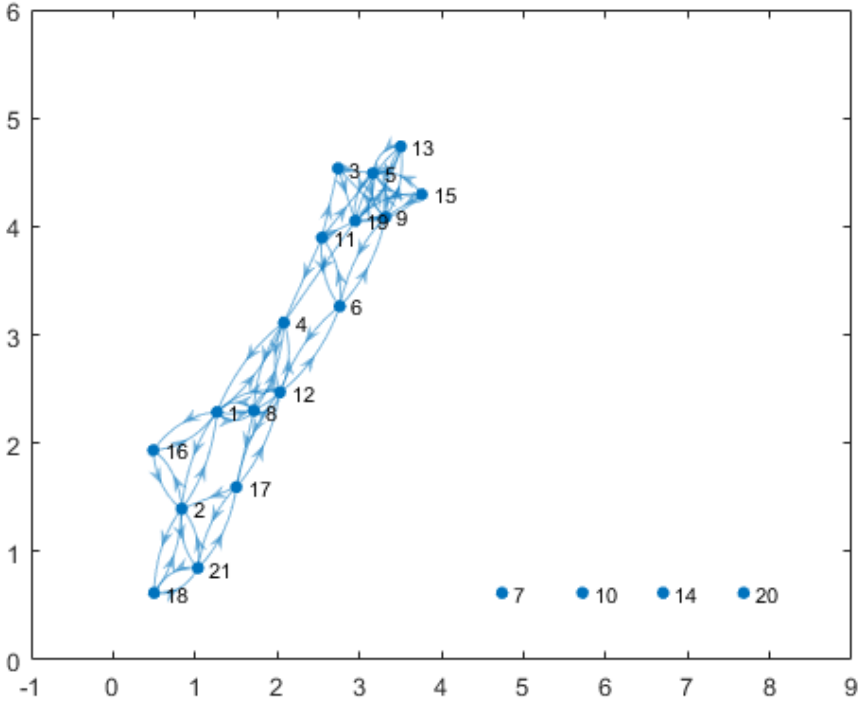
Question

What is the area "under" the curve?

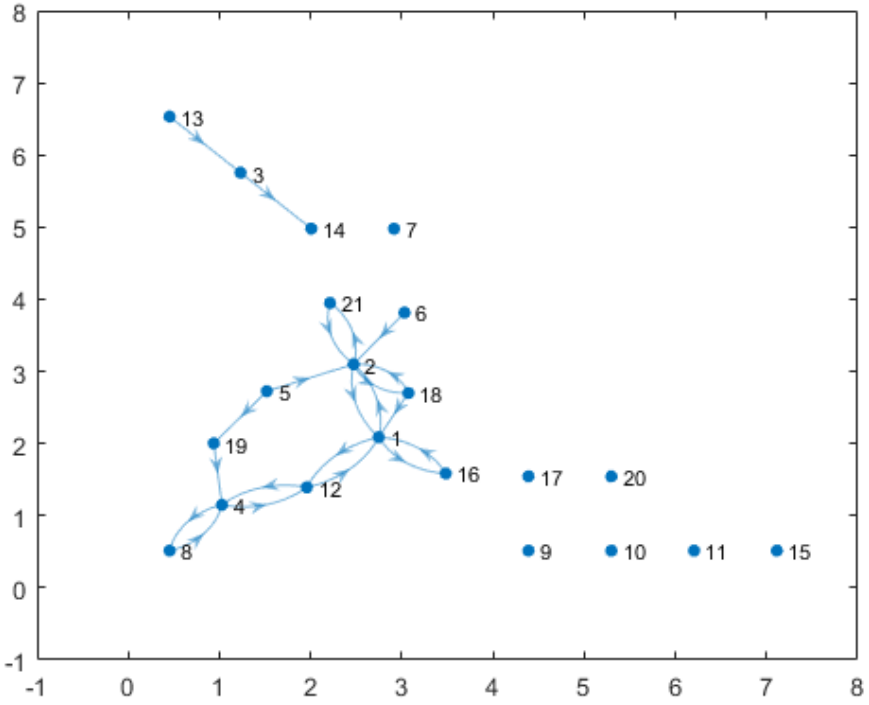
$$\int_0^1 \int_0^{\sqrt{1-x^2}} 1 dx dy$$



Different Perspectives on "Friendship"

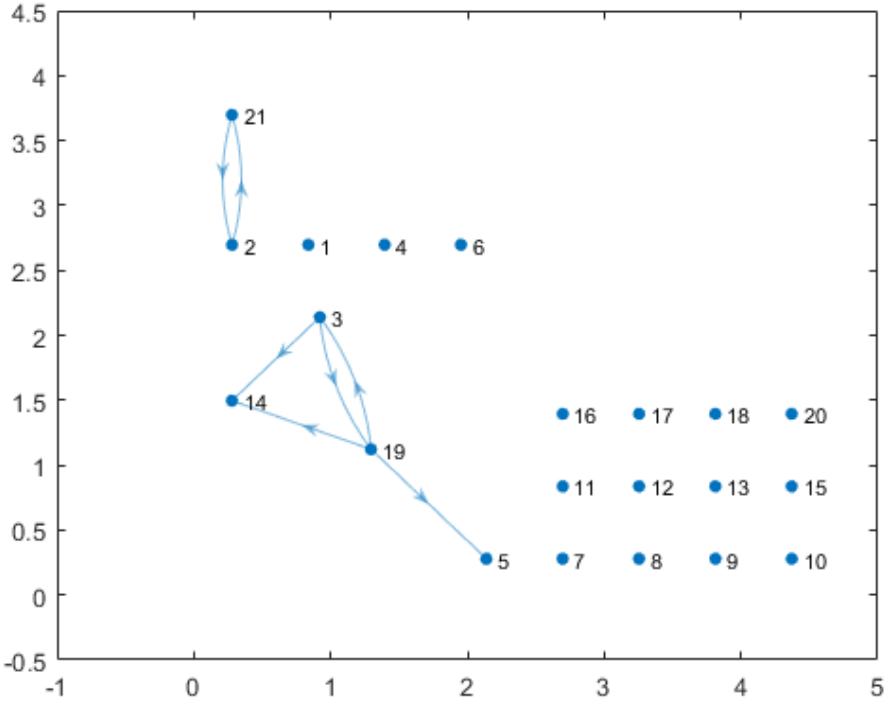


Different Perspectives on "Friendship"



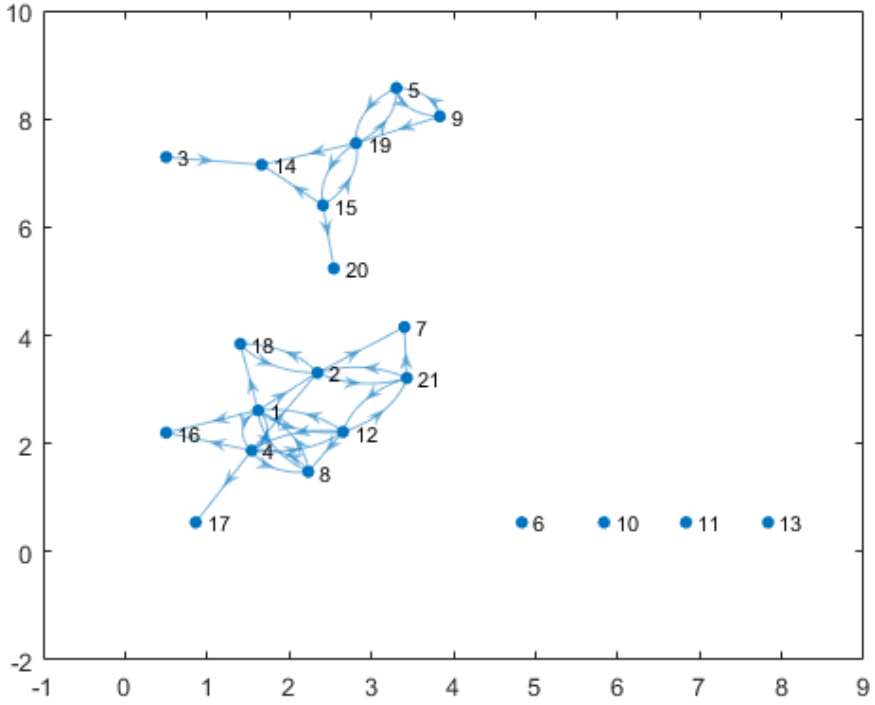
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



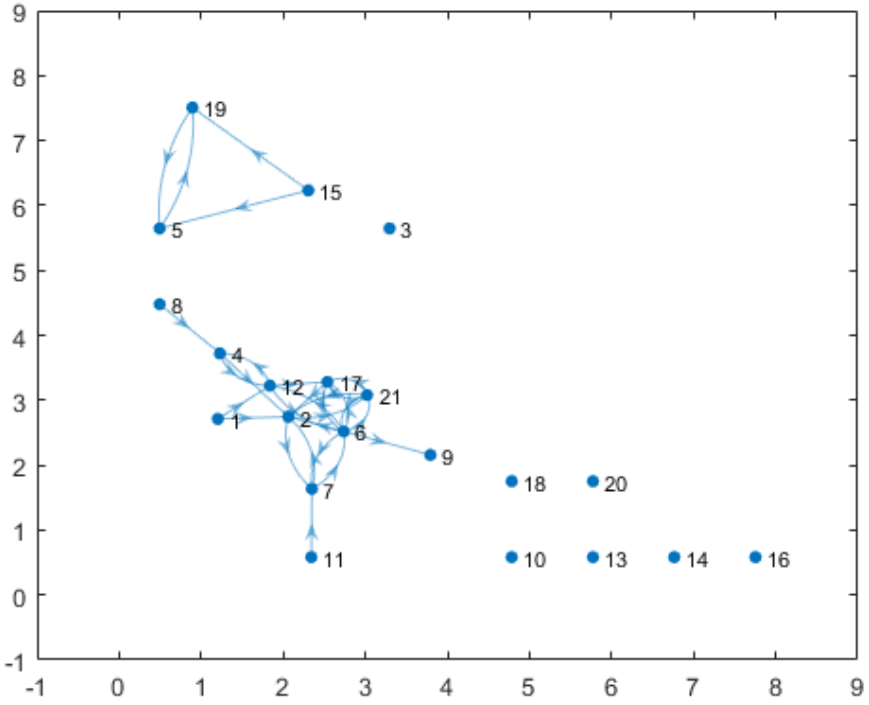
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



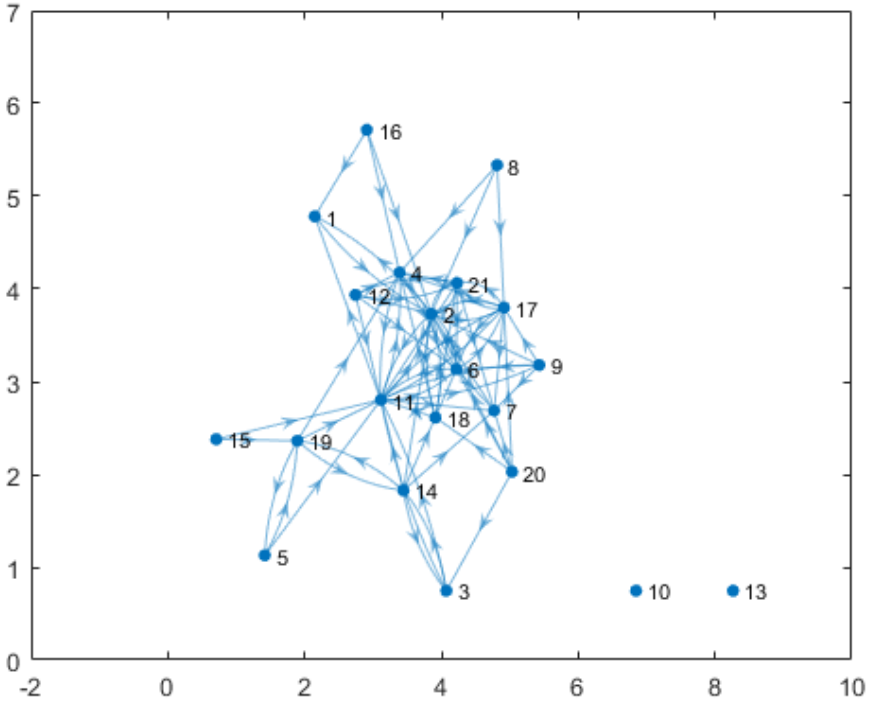
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



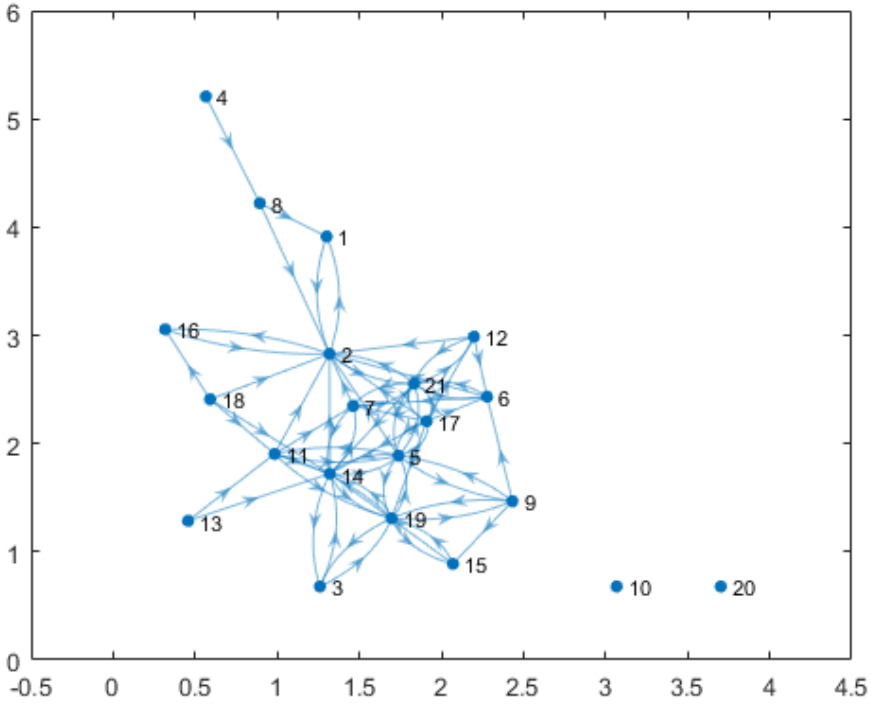
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



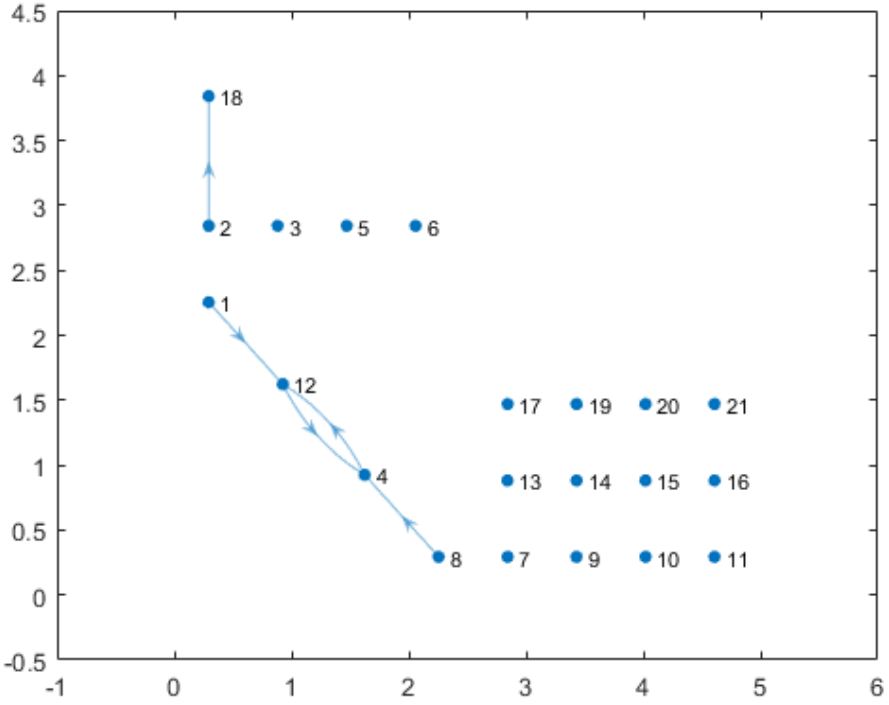
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



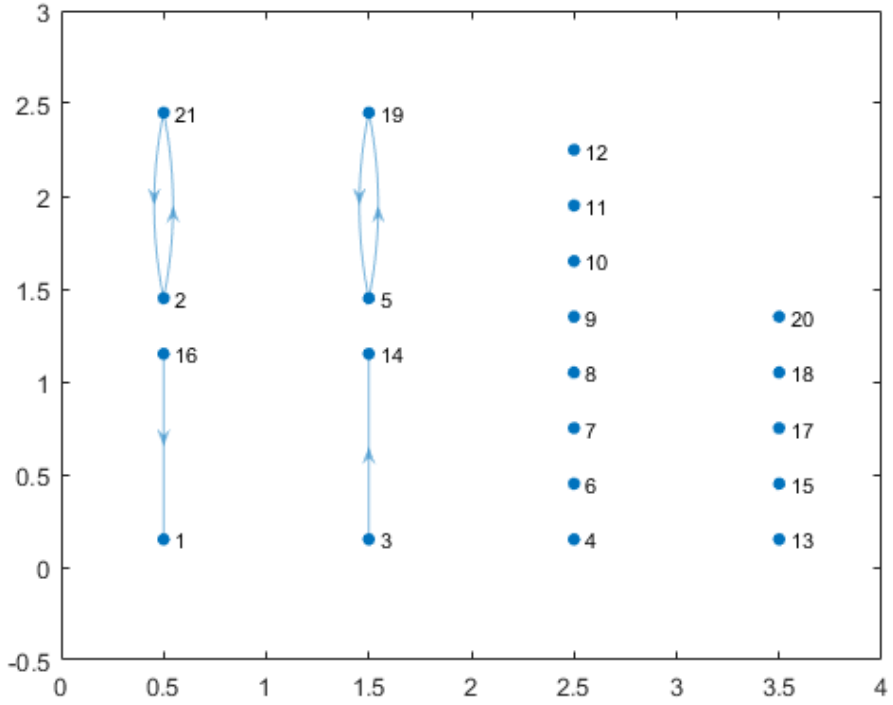
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



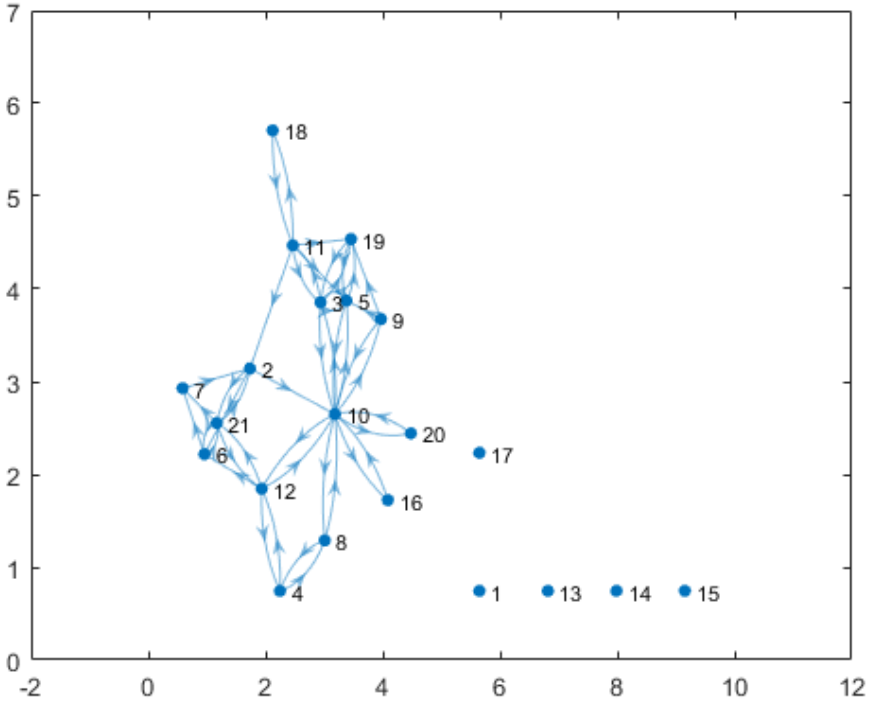
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



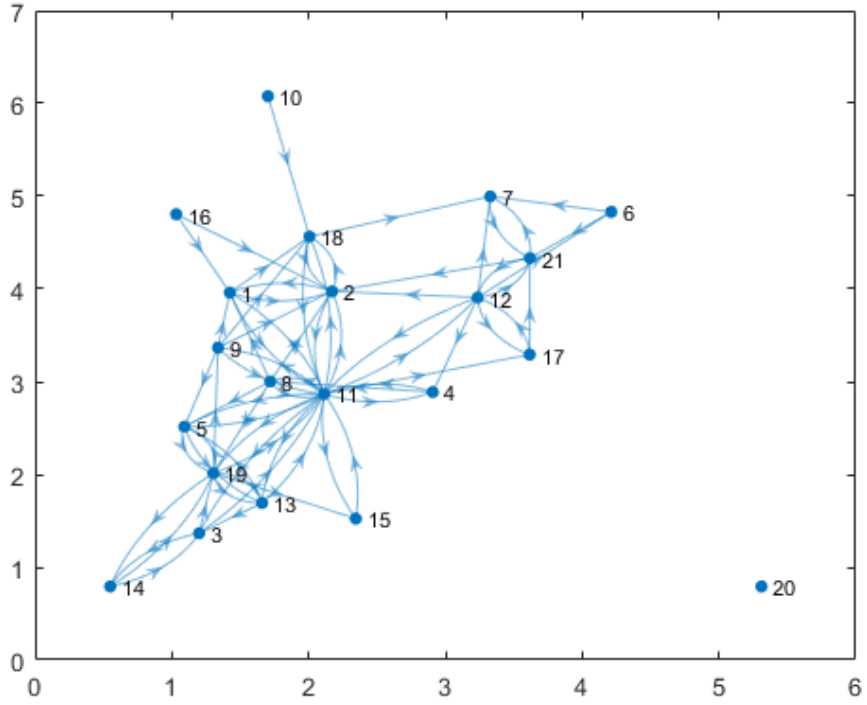
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



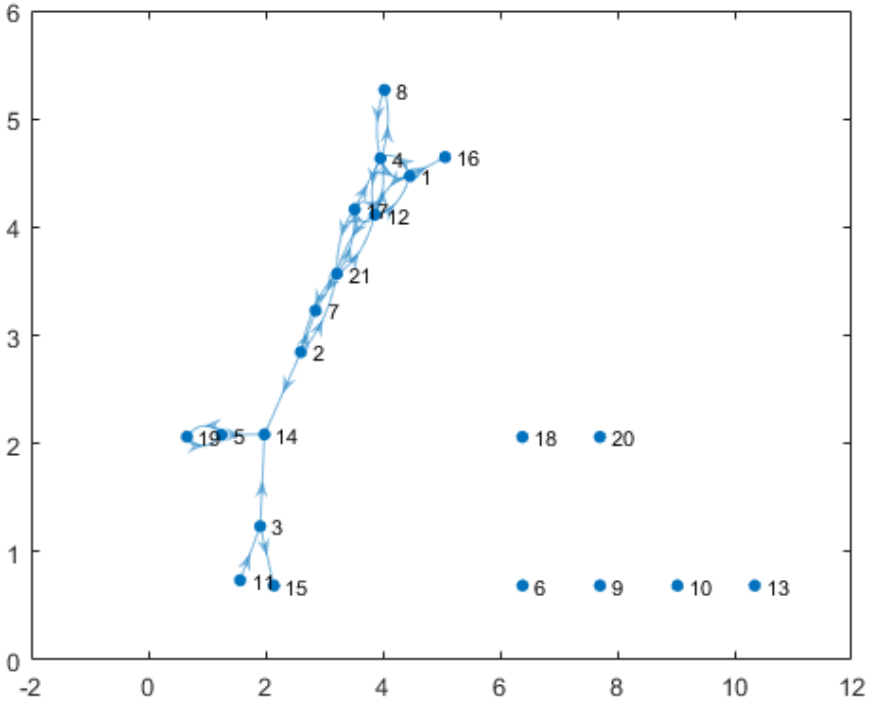
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



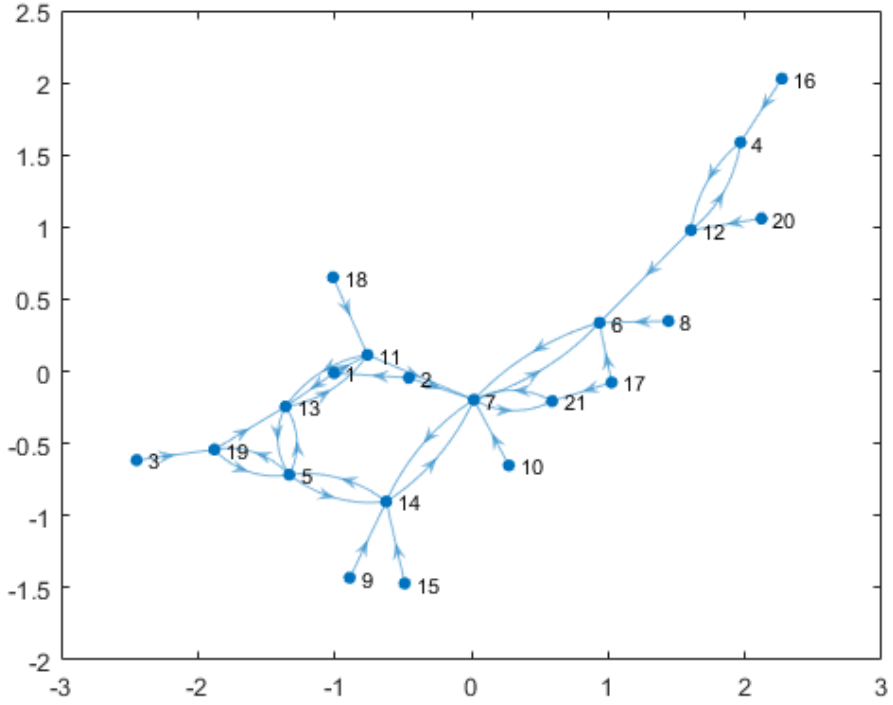
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



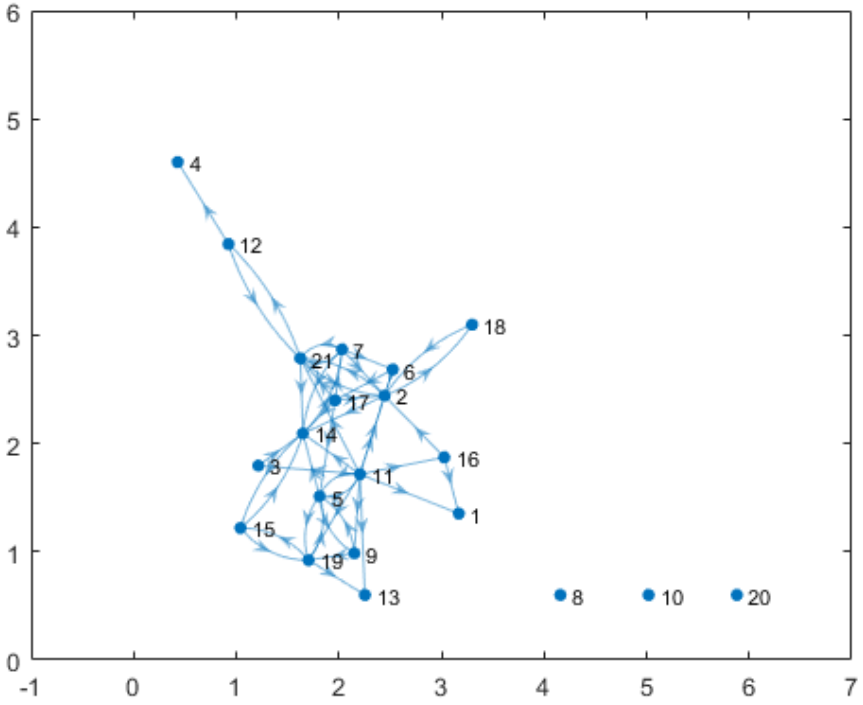
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



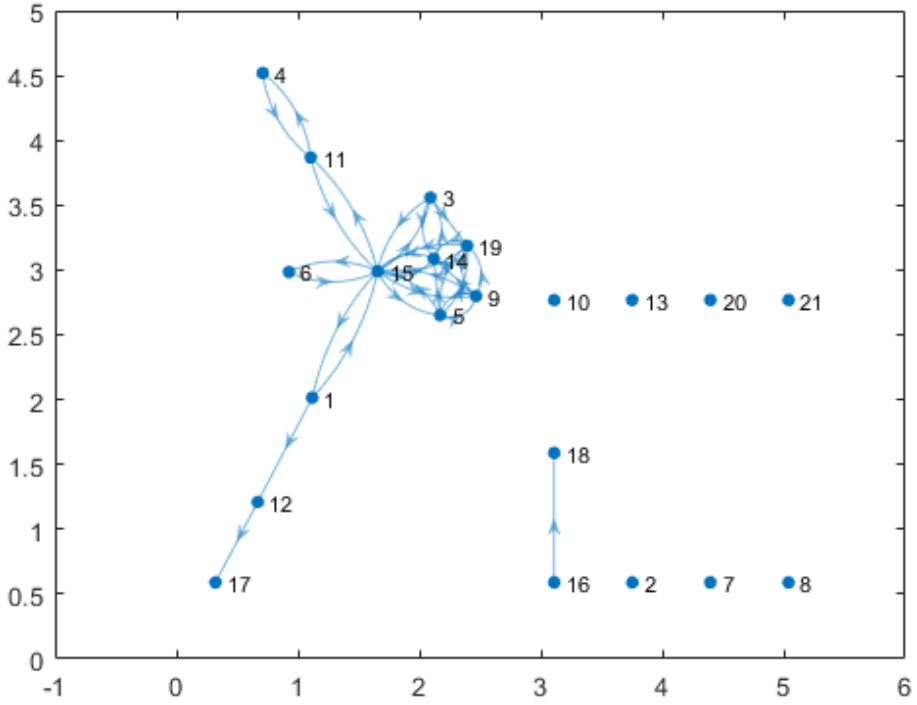
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



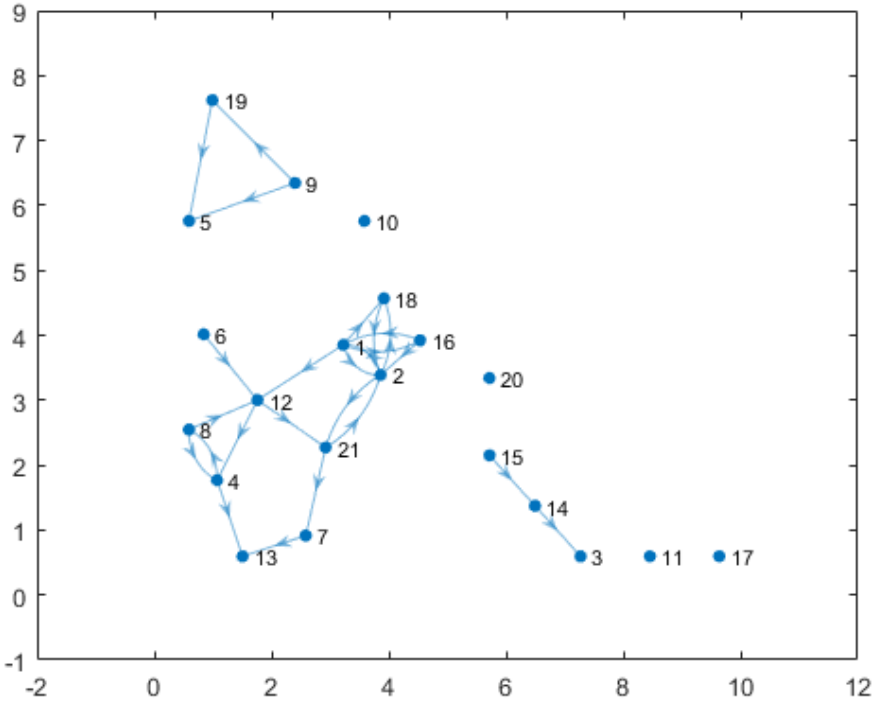
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



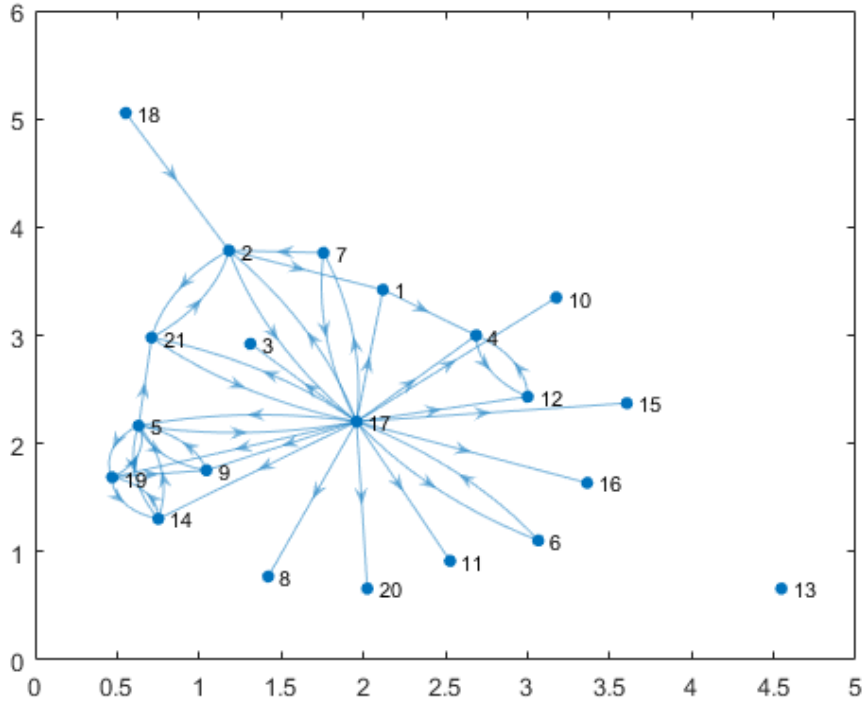
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



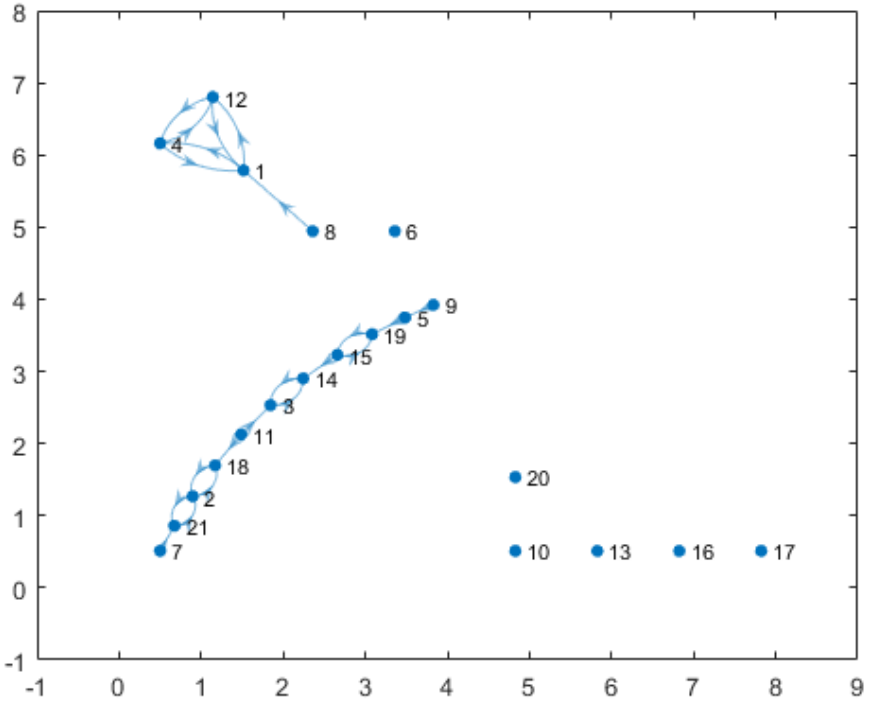
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



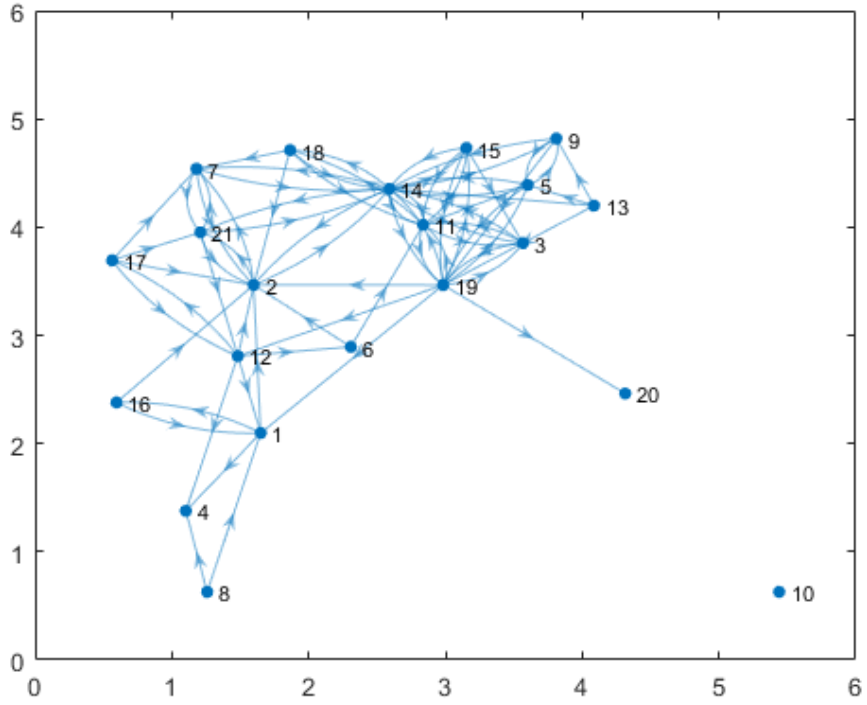
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



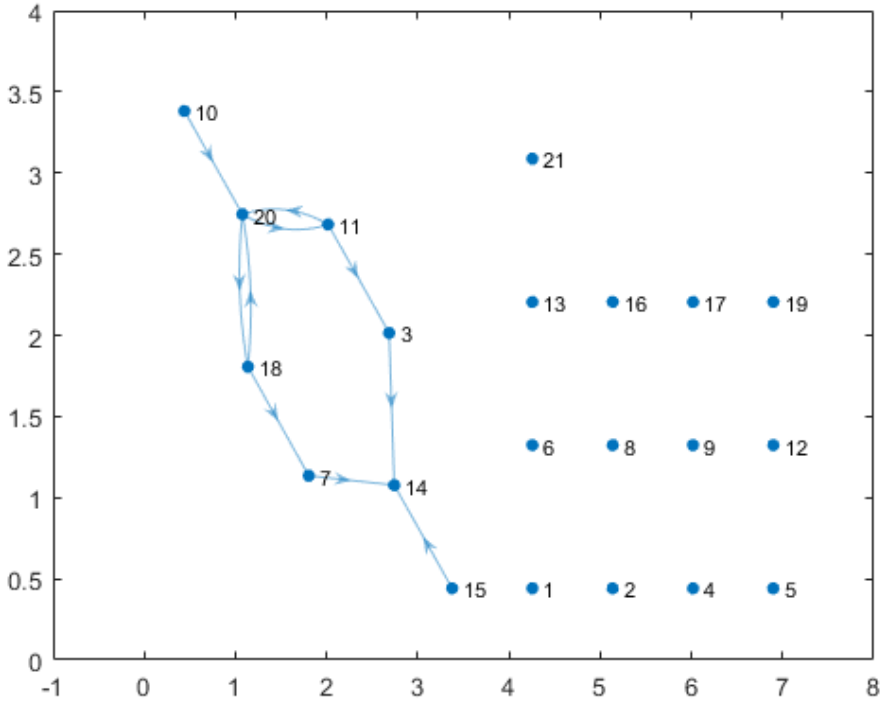
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



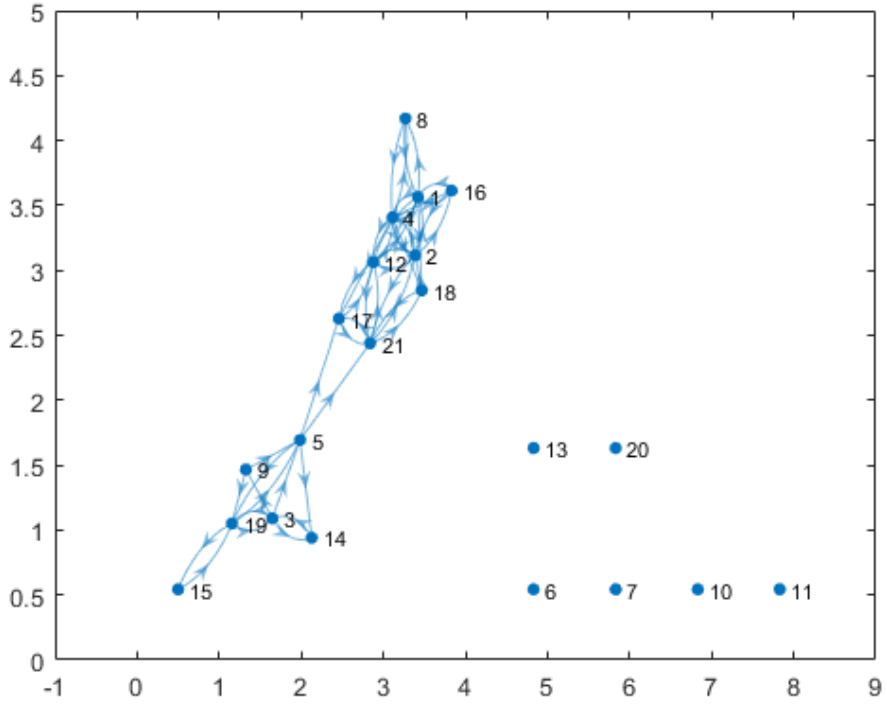
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



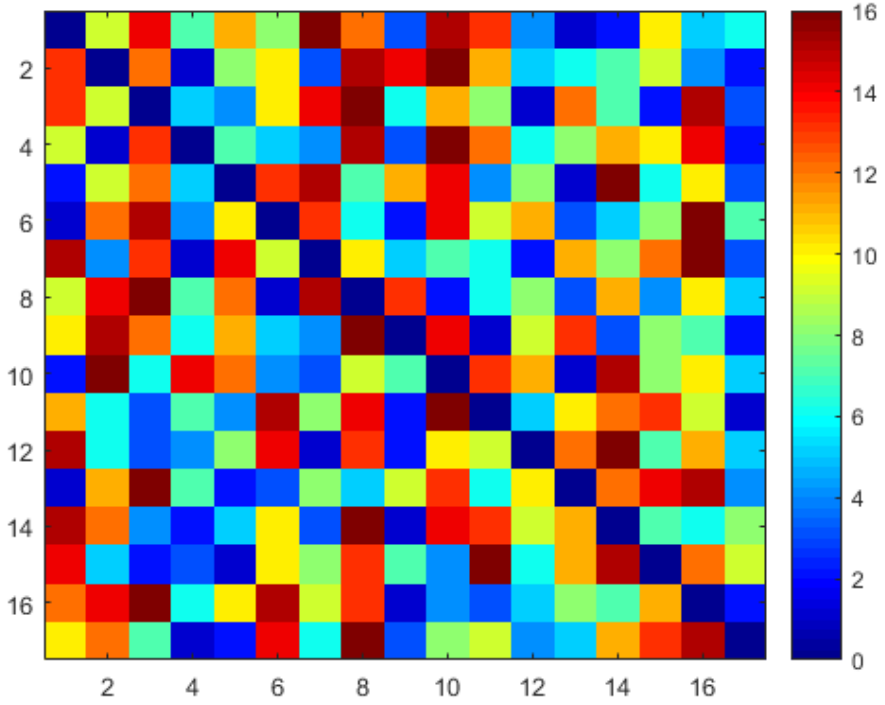
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

Different Perspectives on "Friendship"



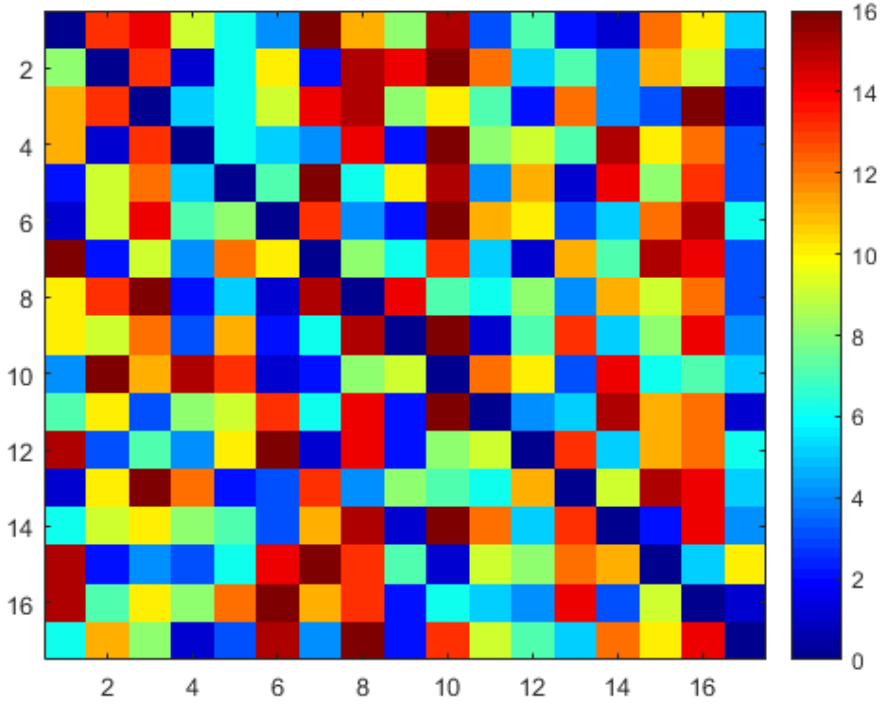
Krackhardt D. (1987). Cognitive social structures. Social Networks, 9, 104-134.

"Friendship" over Time



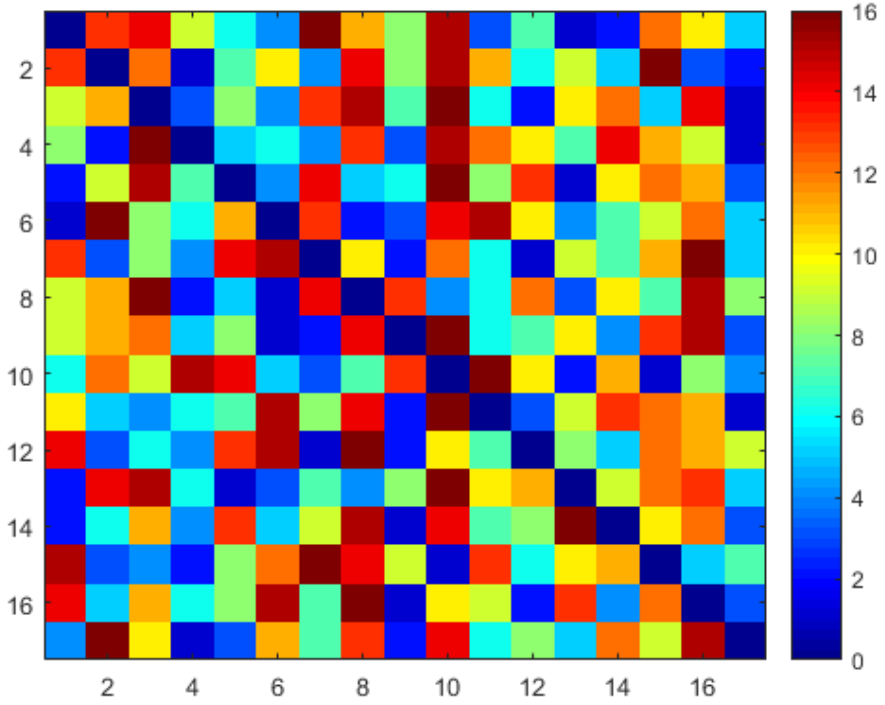
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



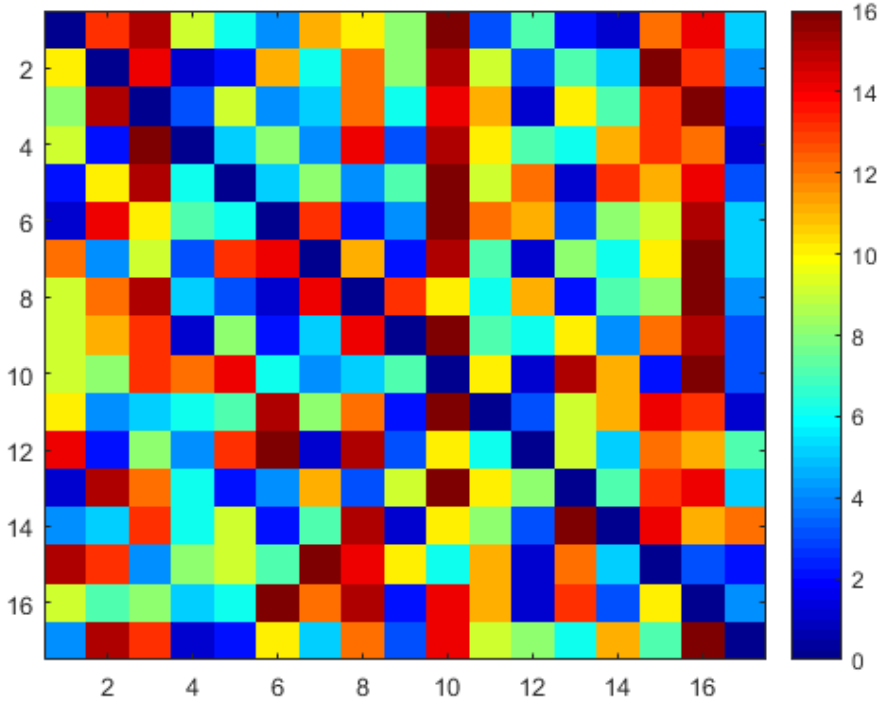
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



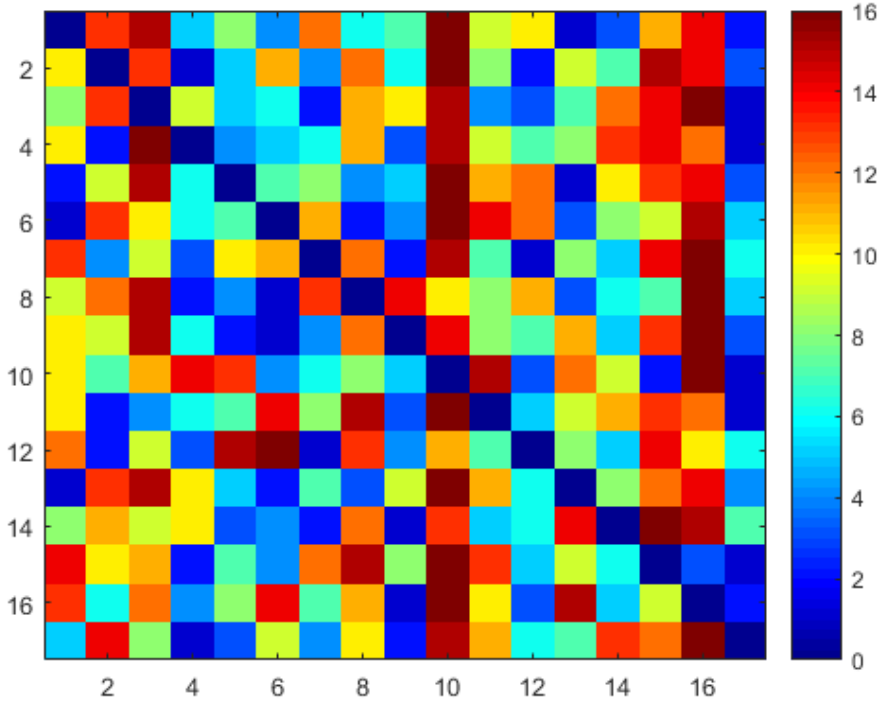
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



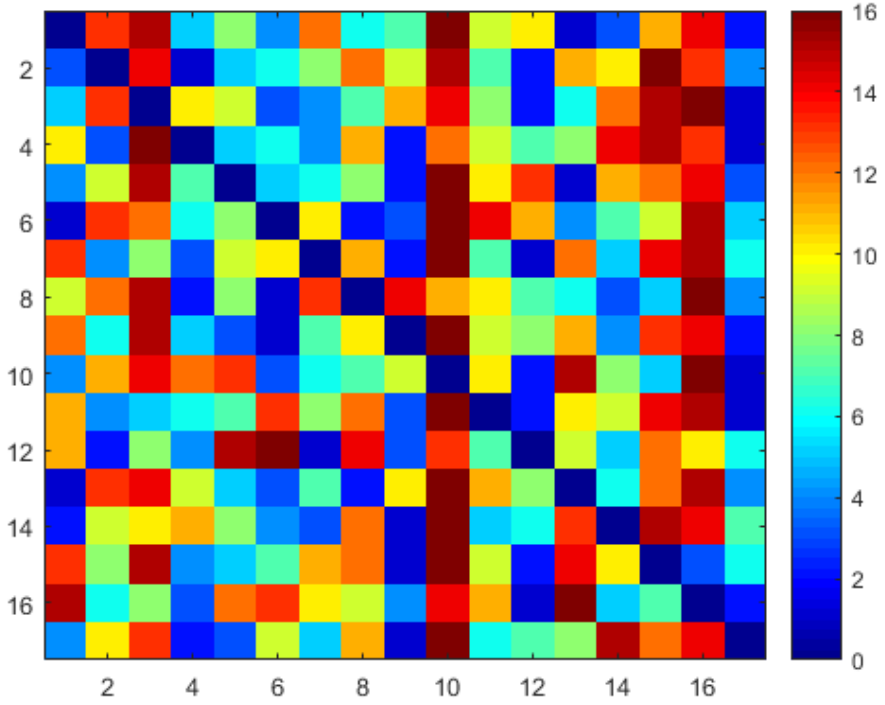
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



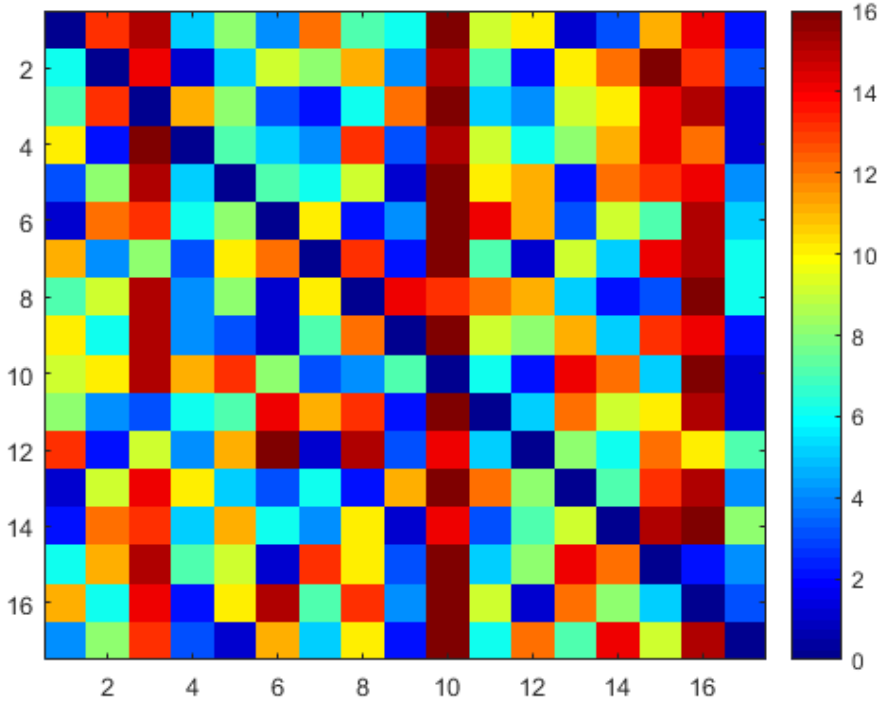
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



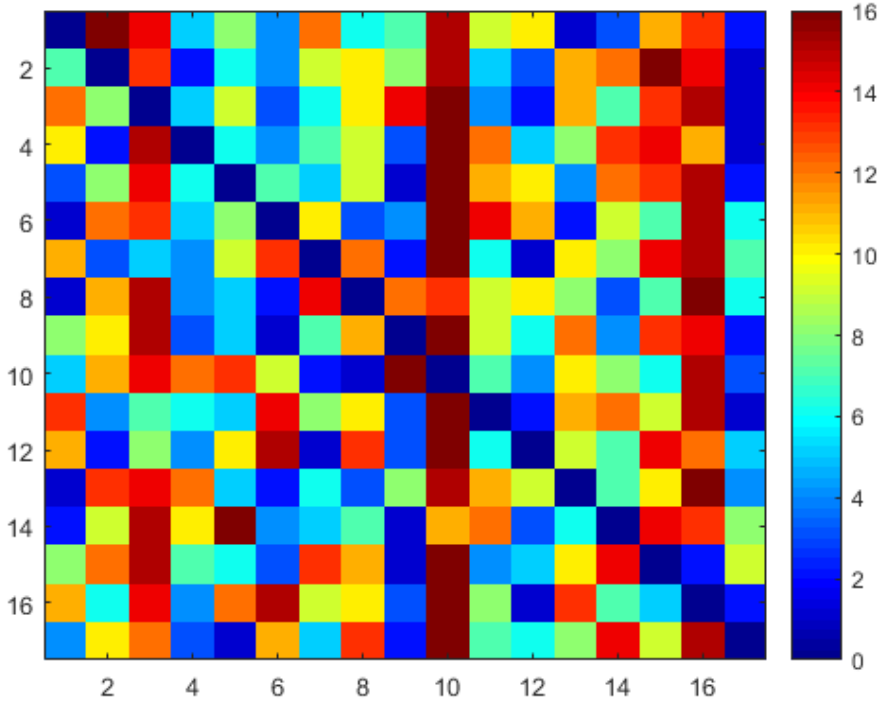
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



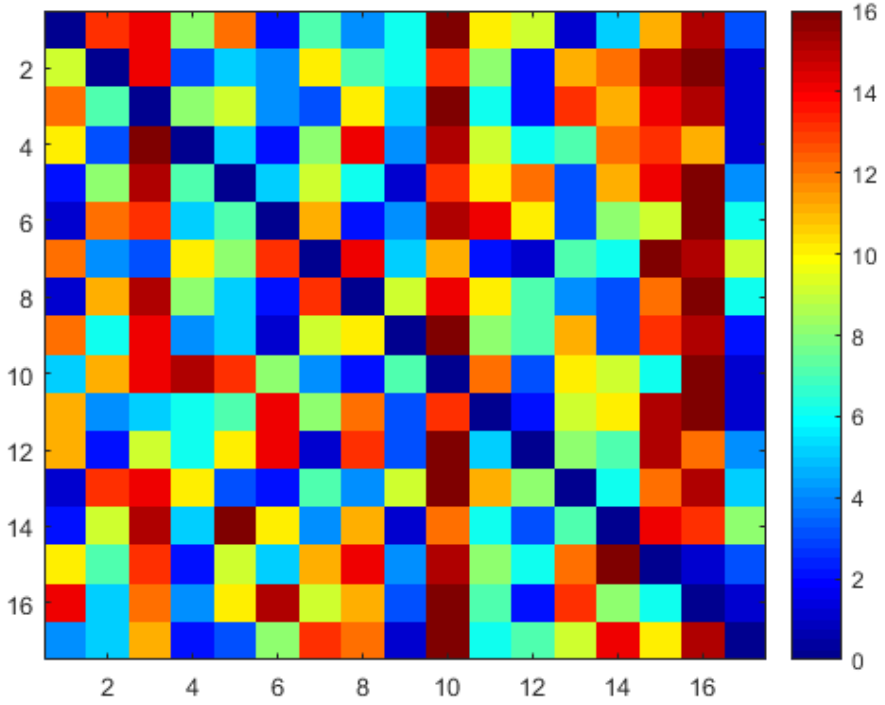
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



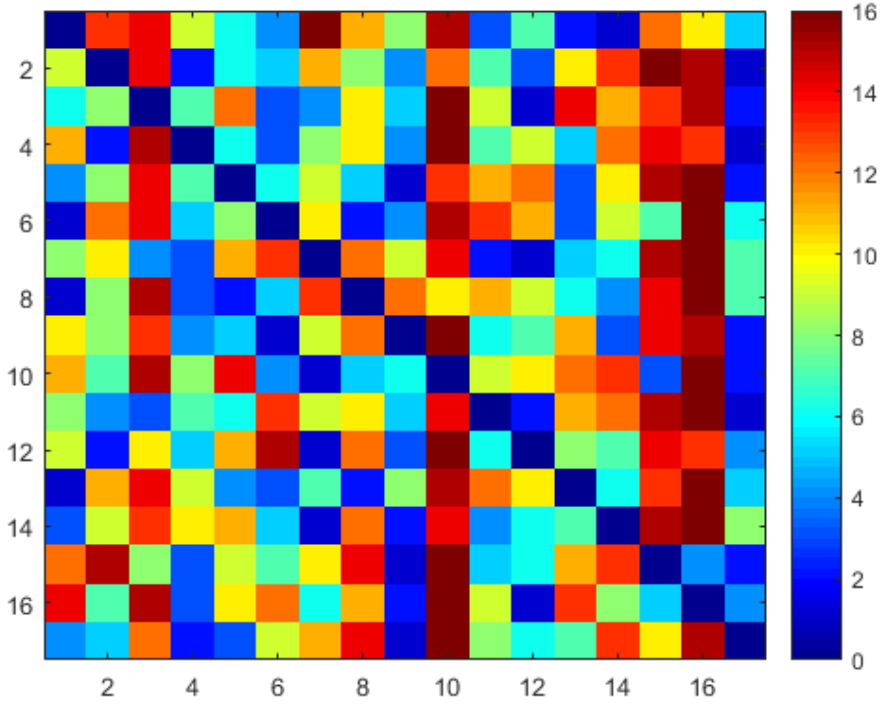
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



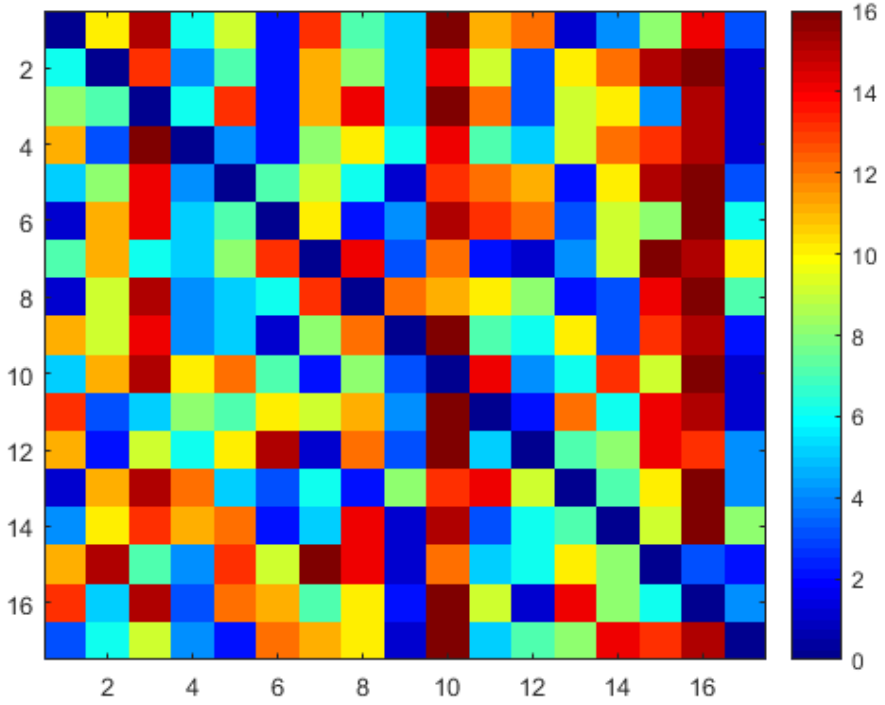
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



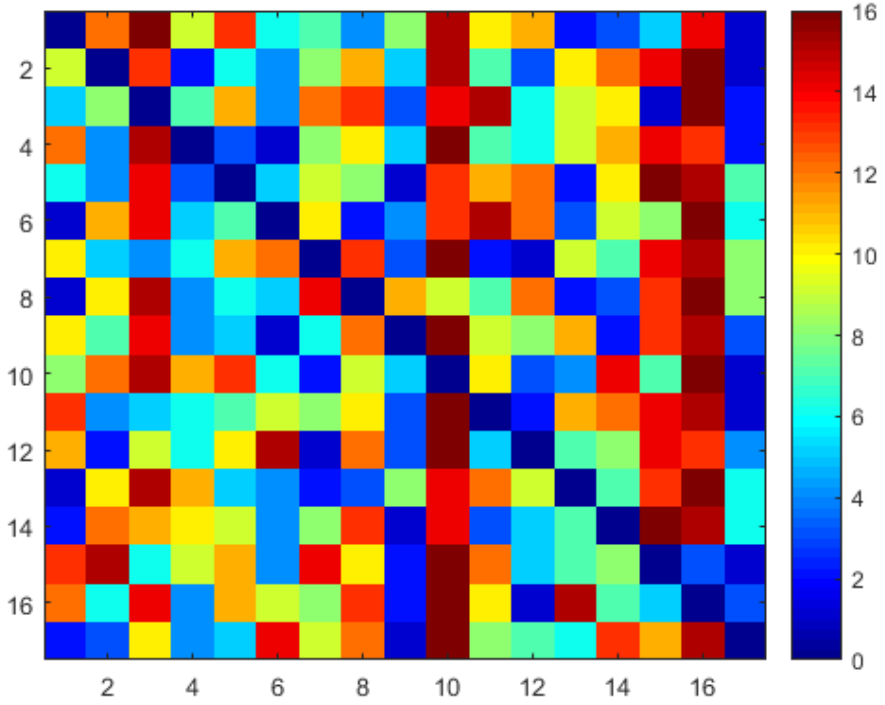
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



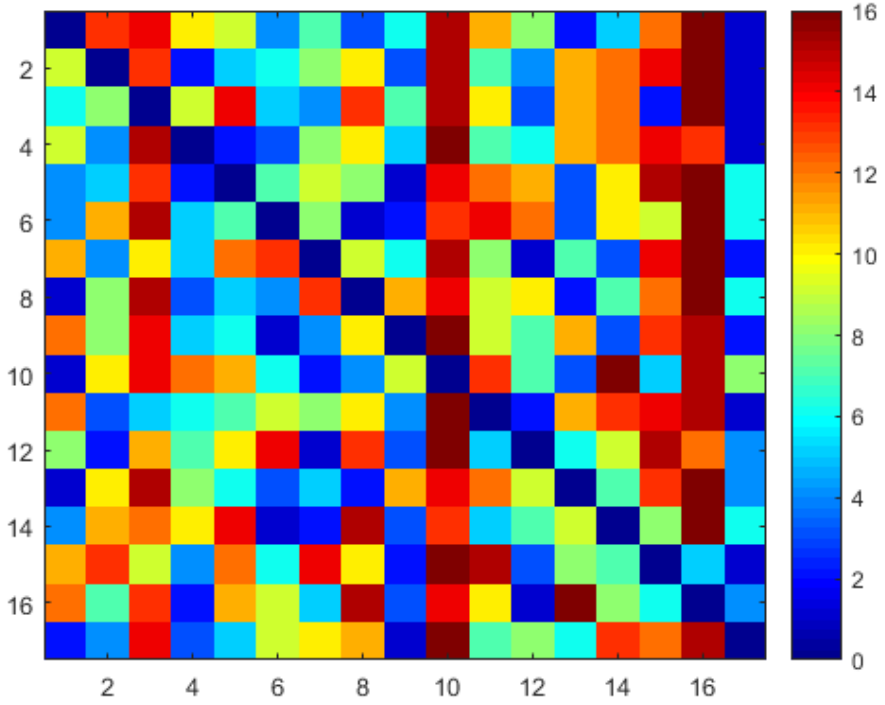
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"Friendship" over Time



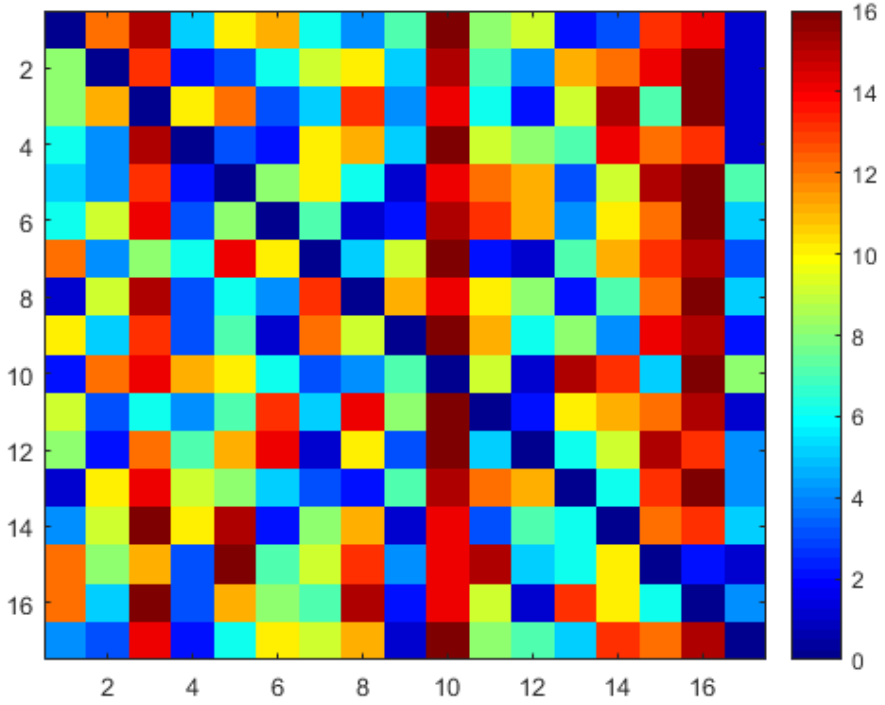
Newcomb T. (1961). The acquaintance process. New York: Holt, Reinhard and Winston.

"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.

Properties of Monte Carlo Methods

- Draw (independent) samples from a random distribution
- Compute some measure for each draw
- Repeat lots and lots of times
- Average/aggregate the derived data

Moderately Ancient

- Buffon's Needle Experiment
- Lord Kelvin (out of a hat)
- Everyone...

Ulam

What is a Markov chain?

Definition (Markov Chain)

A sequence of random variables $X_1, X_2, \ldots,$ is called a Markov Chain if

$$\mathbb{P}(X_n = x_n : X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = x_n : X_{n-1} = x_{n-1}).$$

Examples

- *Snakes & Ladders*
- *Text generation*
- *Walks on graphs (PageRank)*
- *Walks on (families of) graphs (markovchain)*

Markov Formalism

Given a finite state space $X = x_1, x_2, \ldots, x_n$ we can specify a Markov chain over X with transition probabilities $p_{i,j} = \mathbb{P}(X_m = i : X_{m-1} = j)$ and associated transition matrix $P = [p_{i,j}]$.
Desirable adjectives:

- Irreducible: A chain is irreducible if each state is (eventually) reachable from every other state.
- Aperiodic: A chain is aperiodic if for each state, the GCD of the lengths of the loops, starting and ending at that state is equal to 1.
- Steady State Distribution: A distribution π is said to be a steady state of the chain if $\pi = \pi P$.

Key Theorem

If the chain is irreducible and periodic then $\lim_{m\to\infty} P^m = 1\pi$ for a unique π. Even better, if $y_1, y_2, \ldots, y_m$ are samples from π then,

$$\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} f(y_i) = \mathbb{E}[f]$$

The key step of MCMC is to create an irreducible, aperiodic Markov chain whose steady state distribution π is the distribution we are trying to sample from.

What is MCMC?

In our Monte Carlo methods we just required that we sample from our space uniformly but this isn't always easy to do. MCMC gives us a way to sample from a desired pre–defined distribution by forming a related Markov chain (or walk) over our state space, with transition probabilities determined by a multiple of the distribution that we are trying to sample from.

Proportional to a distribution!?!

A common way this arises is when we have a score function or a ranking on our state space and want to draw proportionally to these scores. Given a score $s : X \to \mathbb{R}$ we want to sample from X with probabilities

$$\mathbb{P}(X_i) = \frac{s(X_i)}{\sum_j S_j}$$

When $|X|$ is enormous, we don't want to/can't compute the denominator directly. Also, uniform sampling over–prioritizes low score spaces. This is also an advantage to local methods.

Simplifications

- Discrete probability distribution/state space
- Score function distributions
- Symmetric proposal distributions
- ...

Terminology

Score: A function $s : X \to \mathbb{R}_{\geq 0}$ that determines our target distribution.

Proposal Distribution: A Markov chain Q over X with the property that $Q(x_j : x_i) = Q(x_i : x_j)$.

Terminology

Score: A function $s : X \to \mathbb{R}_{\geq 0}$ that determines our target distribution.

Proposal Distribution: A Markov chain Q over X with the property that $Q(x_j : x_i) = Q(x_i : x_j)$.

Metric: Another function $f : X \to \mathbb{R}$ that is our quantity of interest for the distribution.

Metropolis Procedure

Given that we have a given score function, proposal distribution, metric, and initial graph g_0 we generate new graphs g_n by:

1. Generating $\hat{g}$ according to the proposal distribution $Q(\hat{g} : g_i)$.
2. Compute the acceptance probability: $\alpha = \min\left(1, \frac{s(\hat{g})}{s(g_i)}\right)$
3. Pick a number β uniformly on $[0, 1]$
4. Set

$$g_{i+1} = \begin{cases} \hat{g} & \text{if } \beta < \alpha \\ g_i & \text{otherwise/} \end{cases}$$

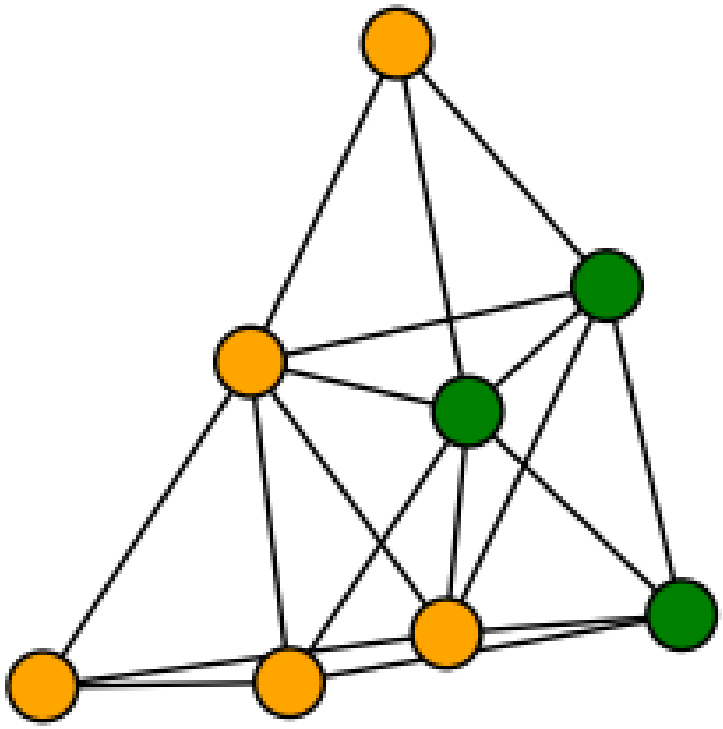
Example:

Score: $s(G)$ is the number of edges within each color

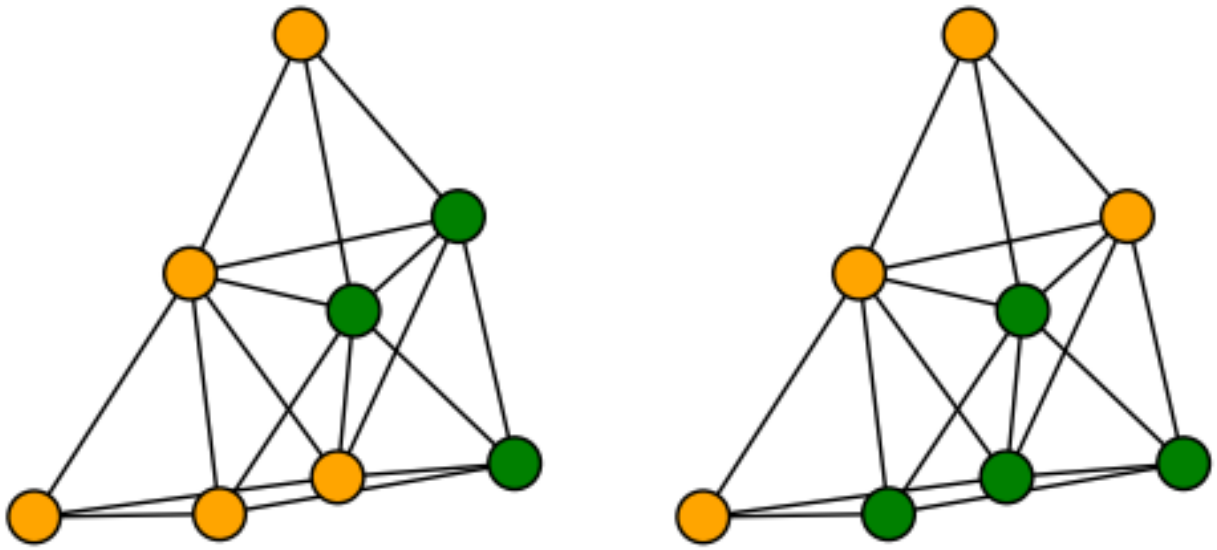
Proposal Distribution: Uniform on partitions with two colors, where each color forms a connected subgraph and the maximum imbalance between the colors is 2 nodes.

Metric: $f(G)$ is the number of edges between two nodes of different colors

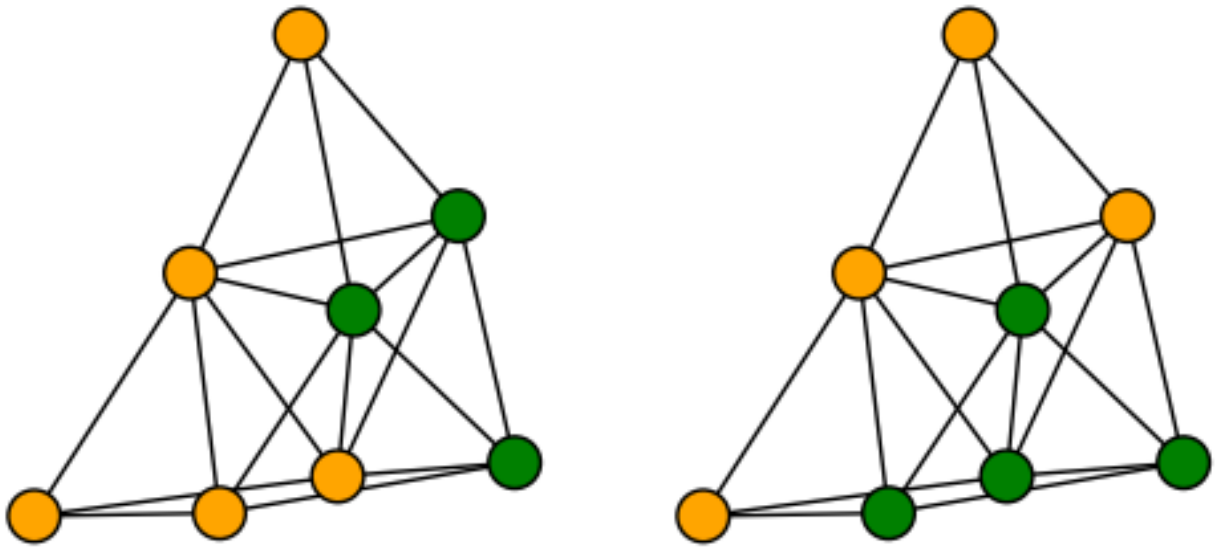
Example: Initial graph



Example: Proposed $\hat{g}$



Example: Proposed $\hat{g}$



Old: 7 edges between yellow nodes and 3 edges between green nodes.
New: 4 edges between yellow nodes and 5 edges between green nodes.

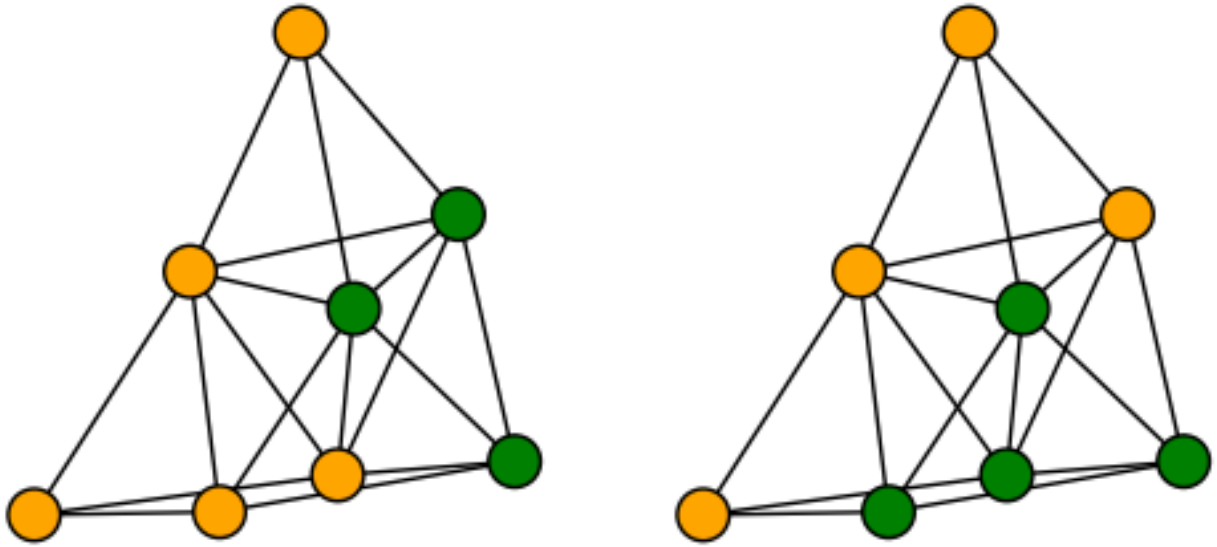
Example: α and β

$$\alpha = \min\left(1, \frac{s(\hat{g})}{s(g_0)}\right) = \min\left(1, \frac{9}{10}\right) = \frac{9}{10}$$

$$\beta = \text{very random} = .31415$$

Success!!

Example: Metric of interest



Old: There are 8 edges between the colors. **New:** There are 6 edges between the colors.

Scores

- Compactness
- Legal Constraints
- Network measures
- ....

Proposal Distributions

- Boundary edge flips
- Uniform
- Node flips
- ...

Metrics

- Compactness
- Legal Constraints
- Network measures
- ....