Recommender Systems
Recommendations in Networks

Professor Robin Burke

Spring 2019

Homework 4

- Due today
- What questions?

Network-based Recommendations

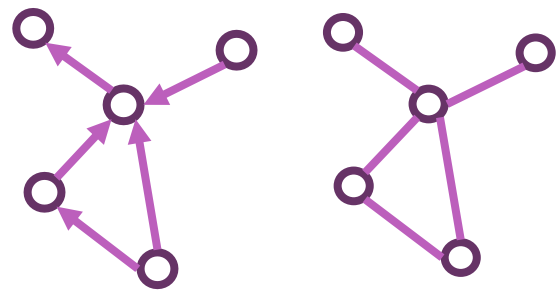
- Covering three approaches
  - Different types of networks
- Reference-based recommendation
  - PageRank and personalized PageRank
- Trust-based recommendation
  - In addition to regular user/item data
  - You have trust links between users
    - Could be friends or other links
    - May also have “distrust” links
- Bipartite User-Item Networks
  - Treat the user-item association as edges in a bipartite network

Network vs Graph

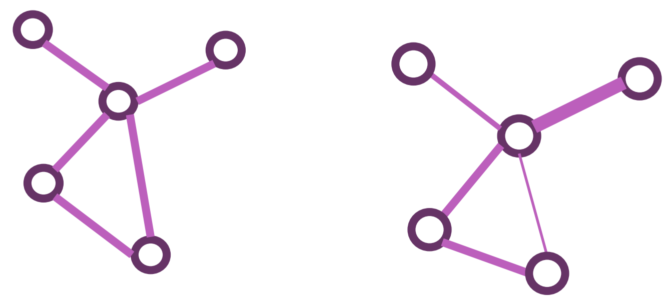
- Networks are real-world phenomena
- Graphs are mathematical and computational representations
- Graphs
  - Nodes / vertices
  - Edges / links / ties etc.

Types of graphs

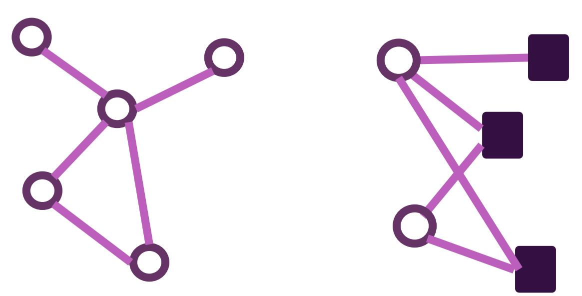
- Directed vs undirected



- Weighted vs unweighted

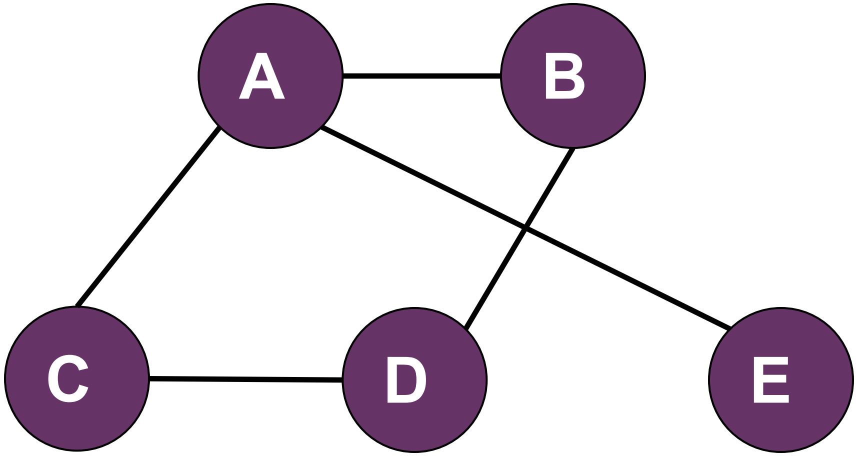


- One-mode vs two-mode (bipartite)



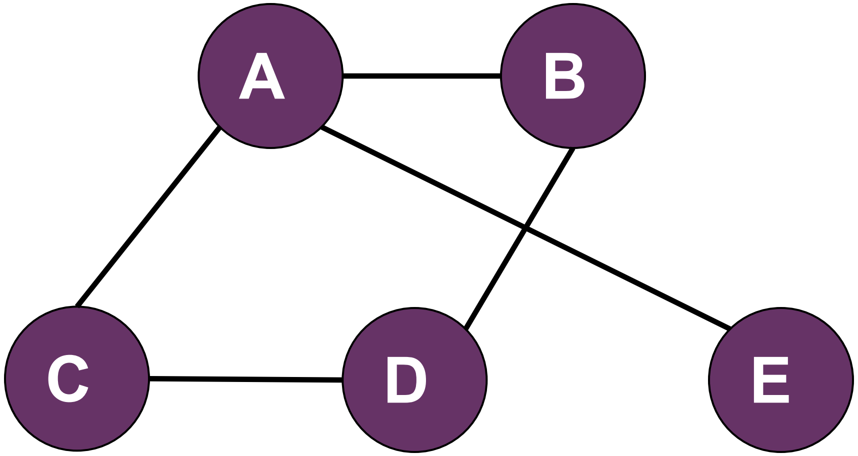
Movement on a graph

- On a graph you can move from node to node
  - via edges
- Might be actual movement
  - messages in a communication network
- Or it might be more implicit
  - social influence



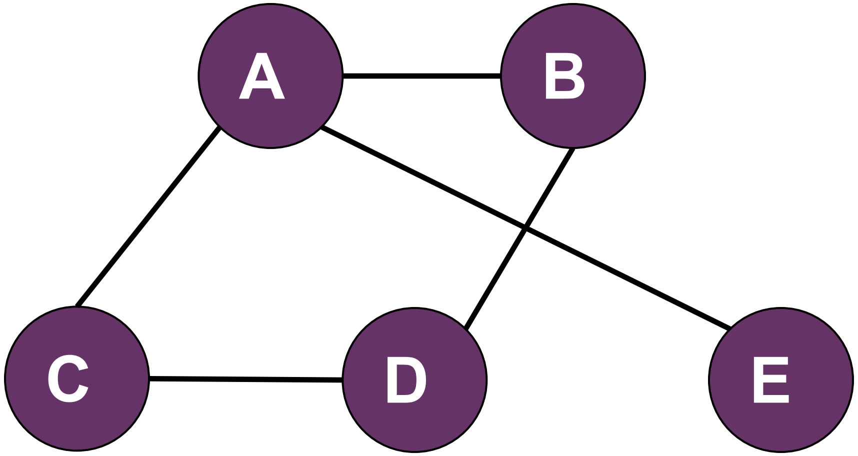
Walk

- Sequence of edges
  - such that the node ending each step is the beginning of the next
- B-A-C-A is a walk…
- Length
  - # of edges



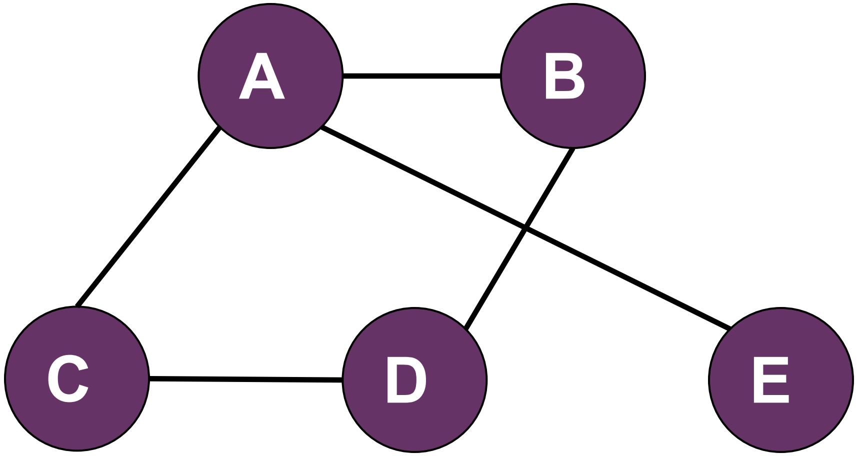
Path

- A walk where no vertex appears more than once
- B-A-C-A is not a path
- B-A-C-D is a path



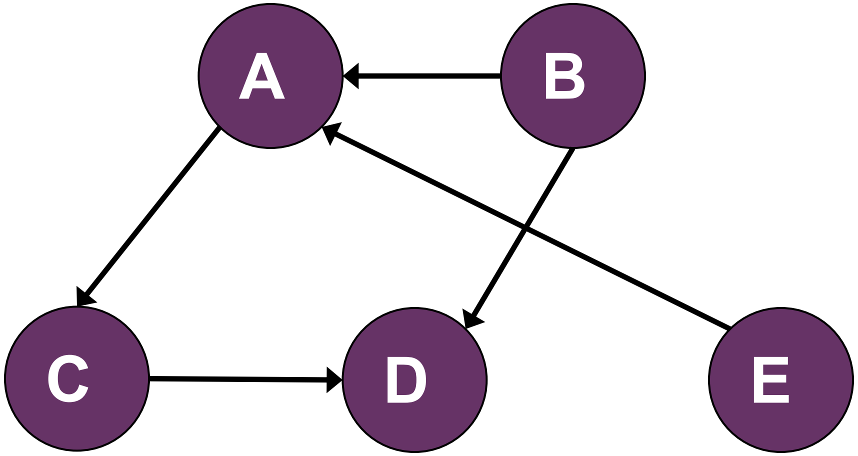
+ Cycle

- A walk where the beginning and endings nodes are repeated
  - but no others
- B-A-B is a cycle
- B-A-C-A is not a cycle
- B-A-C-D-B is a cycle



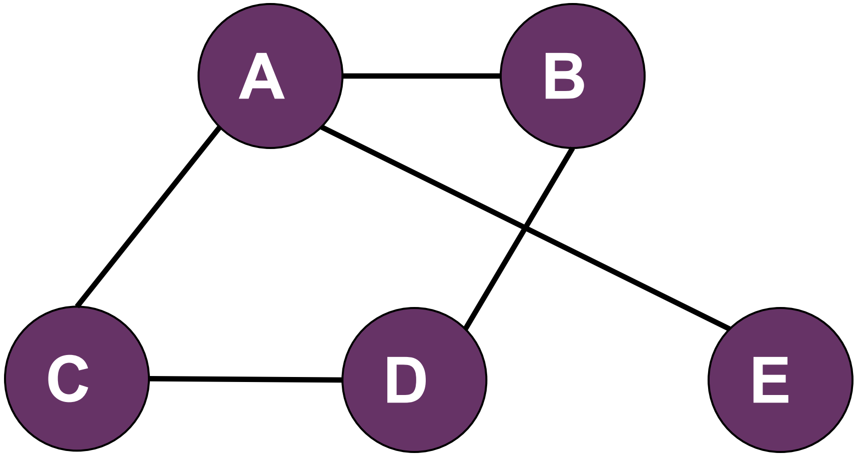
Cyclic / acyclic graphs

- Distinction for directed graphs
- Cyclic graph is one with cycles
- Acyclic has no cycles



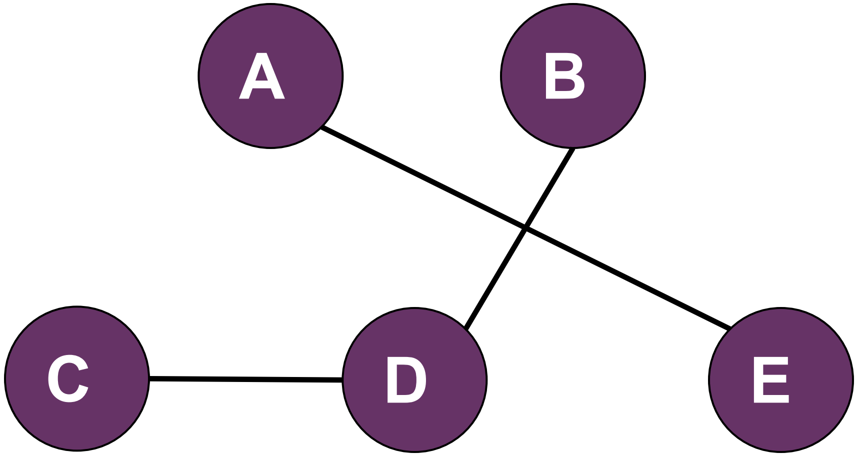
Shortest path / geodesic

- There may be multiple paths from one node to another
- B-A
  - B-A (length 1)
  - B-D-C-A (length 3)
  - B-D-C-D-C-A
    - not a path
- Shortest path aka
  - geodesic



No path

- It is possible that there is no path between two nodes
  - D-E no path
    - length = ∞



Random walk

- A sequence of edges
  - Following the "walk" definition
- Where the edge to traverse is chosen randomly at each node

Alice and Bob live in GraphLand. All the streets are directed edges. Alice tells Bob "I'm going out for a Walk". Bob could:

- A. Call his divorce lawyer because Alice is never coming back to their home node
- B. Ask if he can Cycle along
- C. Ask if she can stop at the GroceryNode before returning
- D. B or C

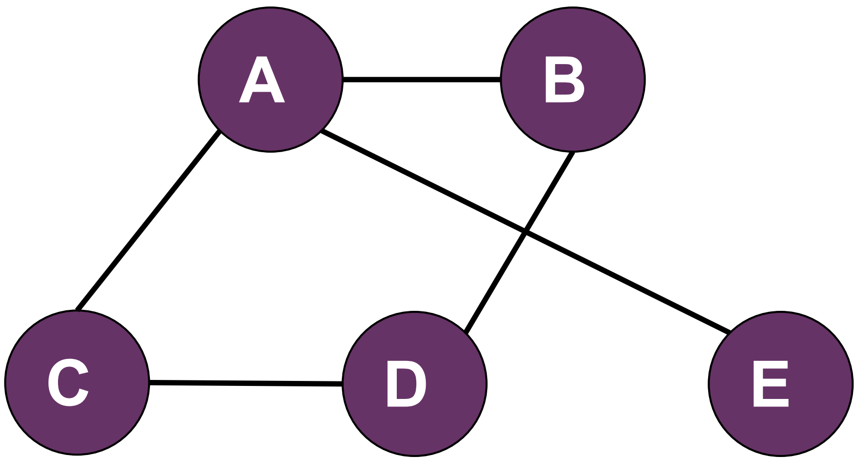
Alice and Bob live in GraphLand. All the streets are directed edges. Bob tells Alice "I'm going out to go out on the garden Path". Alice could:

- A. Call her divorce lawyer because Bob is never coming back to their home node
- B. Ask if she can Walk along
- C. Ask if he can stop at the GroceryNode before returning
- D. B or C

Adjacency Matrix

- Most important graph representation is an adjacency matrix
- Square grid with an entry for each pair of nodes
  - there's a non-zero entry if the nodes are connected
- Matrix cells = edges
  - Can be weighted

Example

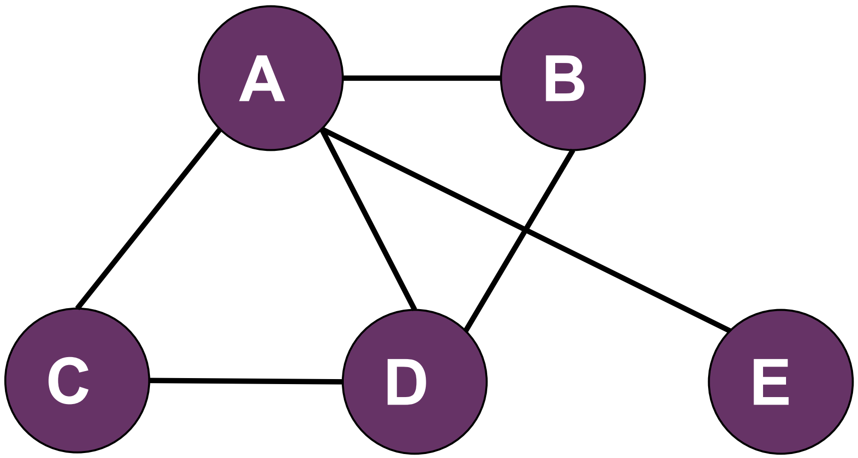


$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Matrix derivation of walks

- Movement in the graph = matrix multiplication
- M x $[1\ 0\ 0\ 0\ 0]^T = [0\ 1\ 1\ 0\ 1]^T$
  - because you can get to B, C, and E from A

+



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

$$M^3 = \begin{bmatrix} 4 & 6 & 6 & 4 & 6 \\ 6 & 2 & 2 & 1 & 5 \\ 6 & 2 & 2 & 1 & 5 \\ 4 & 1 & 1 & 0 & 2 \\ 6 & 5 & 5 & 2 & 4 \end{bmatrix}$$

Diagonal represents…

- All triangles
  - All 3 step walks from n back to itself
  - Divide by 2 for directionality
    - [2 1 1 0 2]
  - Sum and divide by 3 for symmetry
    - 2

$$M^3 = \begin{bmatrix} 4 & 6 & 6 & 4 & 6 \\ 6 & 2 & 2 & 1 & 5 \\ 6 & 2 & 2 & 1 & 5 \\ 4 & 1 & 1 & 0 & 2 \\ 6 & 5 & 5 & 2 & 4 \end{bmatrix}$$

+ Matrix derivation of walks

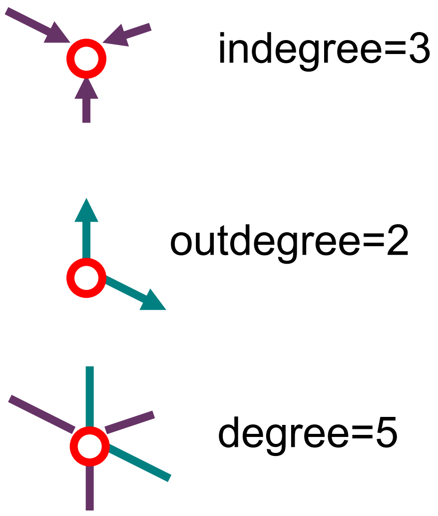
- Means that ideas like
  - Reachability
  - Clustering
  - Random walks
  - Node centrality
- Can be represented as matrix operations
  - Linear algebra
- Like user-item matrices
  - Very sparse

Degree

- Most basic notion of popularity
  - How "connected" is each node?
- High degree nodes have many neighbors
  - more potential for exchange

Types of degree

- in degree
  - how many directed edges are incident (coming in)?
- out degree
  - how many directed edges originate (go out)?
- degree
  - number of edges connected to a node
  - usually applied for undirected networks



Weighted degree

- If edges have weight,
  - might make sense to take this into account
- Example: communication network
  - if weight = # of emails
    - degree = # of contacts
    - weighted degree = total amount of email

In-degree vs out-degree

- These measure very different things
- Which one I care about
  - depends on my question
- Do I want to know about in-flow or out-flow
- In social networks
  - in-degree is often associated with prestige
    - people want to connect to you

I have a weighted directed graph where there is an edge pointing from a user to users whose posts they have commented on. The weight associated with an edge is the number comments made. A user that has high weighted out-degree is one:

- A. Who likes to comment on many people's posts
- B. Whose posts get many comments
- C. Whose posts many users have commented on
- D. Who likes to write many comments

I have a directed graph where there is an edge pointing from a user to users whose posts they have commented on. A user that has high in-degree is one:

- A. Who likes to comment on many people's posts
- B. Whose posts get many comments
- C. Whose posts many users have commented on
- D. Who likes to write many comments