Network Basics

CMSC 498J: Social Media Computing

Department of Computer Science
University of Maryland
Spring 2016

Hadi Amiri
hadi@umd.edu
Lecture Topics

• Graphs as Models of Networks
• Graph Theory
  ▫ Nodes, links, node degree, etc
  ▫ Graph density
  ▫ Complete Graph
  ▫ Graph Connectivity
    • Walks, trails, and paths
  ▫ Reachability
  ▫ Distance and Diameter
  ▫ Adjacency matrix
  ▫ Sub-graphs
  ▫ Graph Types
    • Digraphs, Isomorphic, Bipartite, Multigraphs, Hypergraphs.
Graph Theory

• A graph consists of
  ▫ $\mathbf{N}$: a set of nodes (items, entities, people, etc), and
  ▫ $\mathbf{E}$: a set of links or edges between nodes

• Graph is a way to specify relationships / links amongst a set of nodes.

• We define
  • $\mathbf{N}=|\mathbf{N}| \rightarrow$ size of $\mathbf{N}$
  • $\mathbf{E}=|\mathbf{E}| \rightarrow$ size of $\mathbf{E}$
Graph Theory. Cnt.

• Nodes $i$ and $j$ are adjacent or neighbors if:
  - There is an edge btw them!
    - $i \in \mathbb{N}$
    - $j \in \mathbb{N}$
    - $(i, j) \in E$
Sample Graphs 1.

• “Lives Near” Graph

Sample Graphs 2.

- Brand Proximity Graph

Graphs as Models of Networks

- ARPANET: Early Internet precursor
- December 1970 with 13 nodes
Graphs as Models of Networks - Cnt.

- Only the connectivity matters
  - Could capture distance as weights if needed
Graphs as Models of Networks - Cnt.

- Graph terminology often derived from transportation metaphors
  - E.g. “shortest path”, “flow”, “diameter”
Graphs as Models of Networks - Cnt.

- Abstract graph theory is interesting in itself
- But in network science, items typically represent real-world entities
  - Several examples (from Lecture 1.)
    - Communication networks
      - Companies, telephone wires
    - Social networks
      - People, friendship/contacts
    - Information networks
      - Web sites, hyperlinks
Node Degree $d(i)$

- Given Node $i$, its degree $d(i)$ is:
  - the number of nodes adjacent to it.

Graph Density

• How many edges are possible?
Graph Density- Cnt.

- 5
Graph Density- Cnt.

• $5 + 4$
Graph Density- Cnt.

• $5 + 4 + 3$
Graph Density- Cnt.

- $5 + 4 + 3 + 2$
Graph Density- Cnt.

• $5 + 4 + 3 + 2 + 1$
Graph Density- Cnt.

- \((N-1) + (N-2) + (N-3) + \ldots + 1 = ?\)
Graph Density- Cnt.

- \((N-1) + (N-2) + (N-3) + ... + 1 = N \times (N-1) / 2\)
Graph Density- Cnt.

• Graph Density of a given graph G is determined by:
  ▫ the proportion of all possible edges that are present in the graph, i.e.
  ▫ With N nodes and E edges, graph density is:
    • Number of edges in G / Number of all possible edges in G

\[
\frac{E}{N \times (N - 1)/2}
\]
Graph Density - Cnt.

- Graph Density

\[
\frac{E}{N \times (N - 1)/2}
\]

\[
\frac{6}{[6 \times (6 - 1)/2]} = \frac{6}{15}
\]

What is the density of this graph?

- $N = 16$
- $E = 20$

Complete Graph

- If all edges are present, then all nodes are adjacent (neighbors), and the graph is a Complete Graph.

What is the density of a complete graph?
Graph Connectivity

• Indirect connections between nodes
• We discuss about:
  ▫ Walks
  ▫ Trails
  ▫ Paths
Graph Connectivity - Cnt.

• Walk
  ▫ A sequence of nodes and edges that starts and ends with nodes where each node is incident to the edges following and preceding it.

• Trail
  ▫ A walk is a walk with distinct edges

• Path
  ▫ A walk with distinct nodes & edges.

• The length of a walk, trail, or path is the number of edges in it.
Graph Connectivity- Cnt.

- **Walk**
  - A sequence of nodes and edges that starts and ends with nodes where each node is incident to the edges following and preceding it.
Graph Connectivity - Cnt.

- **Walk**
  - A sequence of nodes and edges that starts and ends with nodes where each node is incident to the edges following and preceding it.

Sample Walk:

\[ W = n_1 \ l_2 \ n_4 \ l_3 \ n_2 \ l_3 \ n_4 \]
• Trail
  ▫ A trail is a walk in which all edges are distinct, although some node(s) may be included more than once.
Graph Connectivity- Cnt.

- **Trail**
  - A trail is a walk in which all edges are distinct, although some node(s) may be included more than once.

Sample Trail:
\[ T = n_4 \ l_3 \ n_2 \ l_4 \ n_3 \ l_5 \ n_4 \ l_2 \ n_1 \]

Graph Connectivity- Cnt.

• **Path**
  - A path is a walk in which all nodes and all edges are distinct.

Graph Connectivity - Cnt.

• Path
  ▫ A path is a walk in which all nodes and all edges are distinct.

Sample Path:
  \[ P = n_1 \, l_2 \, n_4 \, l_3 \, n_2 \]
Graph Connectivity - Cnt.

- Is this a Walk? Trail? Path?
  - Yes, Yes, No
  - We call a closed walk with distinct nodes & edges Cycle!

\[ n_2 \xrightarrow{l_4} n_3 \xrightarrow{l_5} n_4 \xrightarrow{l_3} n_2 \]

Reachability

- If there is a path between nodes $i$ and $j$, then $i$ and $j$ are reachable from each other.
Connected Graph

- A graph is connected if *every pair of its nodes* are reachable from each other
  - i.e. there is a path between them.

Disconnected Graph

How can we make this graph connected?

Connected Graph

and this graph disconnected?

Distance and Diameter

• Distance btw node $i$ and $j$: $d(i,j)$
  ▫ Length of the shorttest path between $i$ and $j$

• Diameter of a graph
  ▫ Diameter of a graph is the maximum value of $d(i,j)$ for all $i$ and $j$

Next session! for now: The path with min number of edges.
What is the distance and diameter of a complete graph?

\[
d(1, 2) = 1 \\
d(1, 3) = 1 \\
d(1, 4) = 2 \\
d(1, 5) = 3 \\
d(2, 3) = 1 \\
d(2, 4) = 1 \\
d(2, 5) = 2 \\
d(3, 4) = 1 \\
d(3, 5) = 2 \\
d(4, 5) = 1
\]

Diameter of graph = \( \max d(i, j) = d(1, 5) = 3 \)
Adjacency Matrix

- Each row or column represents a node!

\[
\begin{pmatrix}
    n_1 & n_2 & n_3 & n_4 & n_5 \\
    n_1 & 0 & 1 & 1 & 0 & 0 \\
    n_2 & 1 & 0 & 1 & 1 & 0 \\
    n_3 & 1 & 1 & 0 & 1 & 0 \\
    n_4 & 0 & 1 & 1 & 0 & 1 \\
    n_5 & 0 & 0 & 0 & 1 & 0 \\
\end{pmatrix}
\]

\[A = A^T\]

Properties of adjacency matrix → next session
Sub-graphs

- Graph $G_s$ is a sub-graph of $G$ if its nodes and edges are a subset of $G$’s nodes and edges respectively.
Sub-graphs- Cnt.

• Graph $G_s$ is a sub-graph of $G$ if its nodes and edges are a subset nodes and edges of $G$ respectively.

Diagram:

- $G$: A graph with nodes a, b, c, d, e, and i, and edges connecting them.
- $G_{s1}$: A sub-graph with nodes a, b, i, and e.
- $G_{s2}$: A sub-graph with nodes a, d, and e.
Graph Types

• We study a few types of graphs:
  ▫ Bipartite graphs
  ▫ Digraphs
  ▫ Multigraphs
  ▫ Hypergraphs
Graph Types- Bipartite Graphs

- A bipartite graph is an undirected graph in which:
  - nodes can be partitioned into two (disjoint) sets $N_1$ and $N_2$ such that:
    - $(u, v) \in E$ implies either $u \in N_1$ and $v \in N_2$ or vice versa.
  - In other words, all edges go between the two sets $N_1$ and $N_2$ but are not allowed within $N_1$ and $N_2$.

$N_1=$movies $N_2=$actors

$N_1=$\{a,b,c,d\}$N_2=$\{x,y,z\}
Graph Types- Digraphs

• Digraphs or Directed Graphs  
  ▪ Edges are directed
• Adjacency:  
  ▪ There is a direct edge btw nodes!
    • $i \in N$
    • $j \in N$
    • $(i,j) \in E$
Graph Types- Digraphs- Cnt.

• Node Indegree and Outdegree
  ▫ Indegree
    • The indegree of a node, $d_I(i)$, is the number of nodes that links $i$,
  ▫ Outdegree
    • The outdegree of a node, $d_O(i)$, is the number of nodes that are linked by $i$,

• Indegree: number of edges terminating at $i$.
• Outdegree: number of edges originating at $i$. 
Graph Types - Digraphs - Cnt.

\[ d_O(n_i) = \sum_{j=1}^{n} A_{ij} \]

\[ d_I(n_j) = \sum_{i=1}^{n} A_{ij} \]

\[
A = \begin{pmatrix}
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
2 \\
2 \\
1 \\
1 \\
1 \\
1 \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
0 & 3 & 1 & 0 & 2 & 2 \\
\end{pmatrix}
\]

\[ A \neq A^T \]
Graph Types - Digraphs - Cnt.

- Density of Digraph:
  - Number of all possible edges in Digraph?
    - \( N \times (N-1) \)

\[
\frac{E}{N \times (N - 1)}
\]
Graph Types- Digraphs- Cnt.

- Connectivity
  - Walks
  - Trails
  - Paths

- The same as before just links are directed!
Graph Types- Multigraphs

• A Multigraph (or multivariate graph) \( G \) consists of:
  ▫ a set of nodes, \textit{and}
  ▫ two or more sets of edges, \( E^+ = \{E_1, E_2, \ldots, E_r\} \), \( r \) is the number of sets of edges
Multigraph 1.

Source: the geography of transport systems http://www.people.hofstra.edu/geotrans/
Multigraph 2.
Graph Types- Multigraphs- Cnt.

• Each $E_i$ indicated one type of relationship, e.g.:
  ▫ $E_1$: lives near relationship
  ▫ $E_2$: friends at the beginning of the year
  ▫ $E_3$: friends at the end of the year
Graph Types - Multigraphs -Cnt.

• Number of edges btw any two nodes in a multigraph?
  - \( E^+ = \{E_1, E_2, \ldots, E_r\} \), \( r \) is the number of sets of edges
    - Undirected multigraph
      - \([0, r]\)
    - Directed multigraph
      - \([0, 2^r]\)
Graph Types- Hypergraphs

- A hypergraph is a graph in which an edge can connect any number of nodes.
- In a hypergraph, $E$ is a set of non-empty subsets of $N$ called hyperedges.
Graph Types - Hypergraphs - Cnt.

- A hypergraph is a graph in which an edge can connect any number of nodes.
- In a hypergraph, $E$ is a set of non-empty subsets of $N$ called hyperedges.

$N = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$

$E = \{e_1, e_2, e_3, e_4\} =$

$\{\{v_1, v_2, v_3\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}, \{v_4\}\}$
Weighted Graphs

• Edges may carry additional information
  ▫ Tie strength → how good are two nodes as friends?
  ▫ Distance → how long is the distance btw two cities?
  ▫ Delay → how long does the transmission take btw two cities?
  ▫ Signs → two nodes are friends or enemies?

• Such graphs are called weighted or signed graphs and we will study them later.
Isomorphic Graphs

• Isomorphic
  ▫ Two graphs are isomorphic if:
    • there is a one-to-one mapping btw their nodes that preserves adjacency!

Questions?
Announcements

• Please register on Piazza
  ▫ for important announcements
  ▫ forum discussions

• Hadi's office hours (UPDATED):
  ▫ 12:15-1:30pm, or by appointment.

• HWs due time and date (UPDATED):
  ▫ 11:00am on Tuesday classes
    • late within 2:30 hour (i.e. until the end of Hadi's office hours): 10% reduction in grade;
    • after that: zero mark.
Reading

• Ch.02 Graphs [NCM]