Fundamentals

Center for Public Health Systems Science
Brown School
Washington University in St. Louis

Overview

- Describe three main notation systems for describing social networks
- Introduce basic concepts in graph theory
  - Important building blocks for subsequent concepts and methods

Nodes and edges

STARTING DEFINITIONS

Actor/Node/Vertex

- Social entity that my be linked to other social entities
- Not limited to persons, may be organizations, political entities, animals, etc.
- ‘Vertex’ comes from graph theory; ‘Node’ used more often in social network analysis

Relation/Link/Edge/Arc

- Any type of social tie that links two actors
- Two types of relations
  - Directional – indicates transmission or flow from one node to another
  - Non-directional – indicates the existence of connection or relationship between two nodes

Multiple ways to describe networks

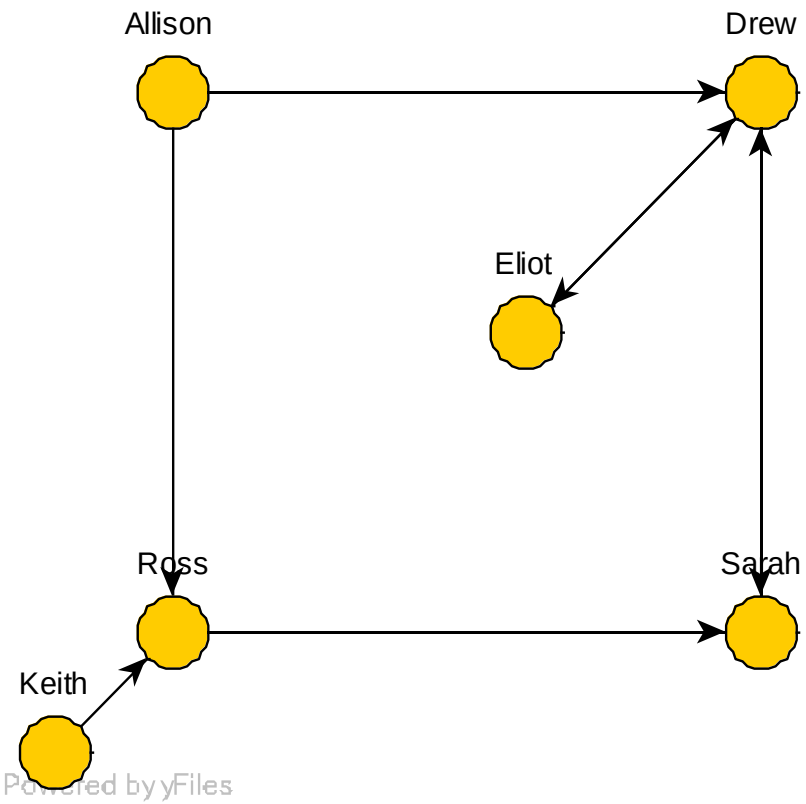
NOTATION

Multiple ways to describe networks

- Interrelated--All approaches describe the same network
  - Graph theoretic
  - Sociometric
  - Algebra/Set theory
  - Edge list
- Some approaches are more conducive to certain types of operations on the networks than others
  - Calculations
  - Storage
  - Transformations

Graph theoretic

- Basic way to present actors and relations
- A graph consists of nodes joined by lines
- Example (W&F, Fig. 3.1)

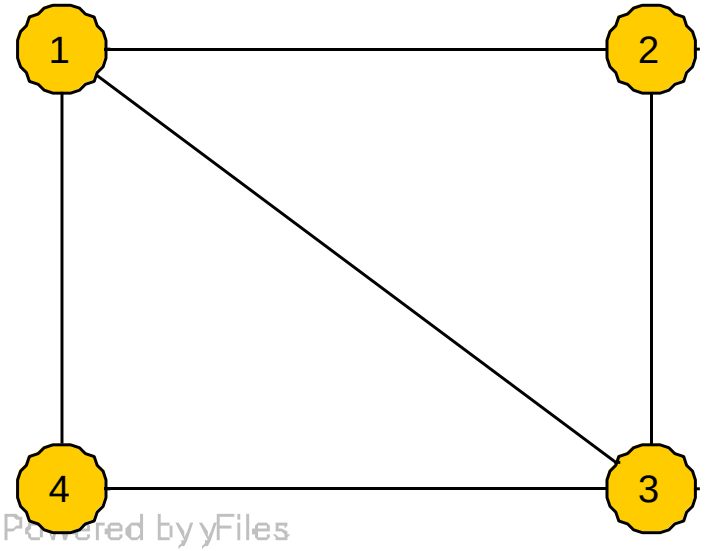


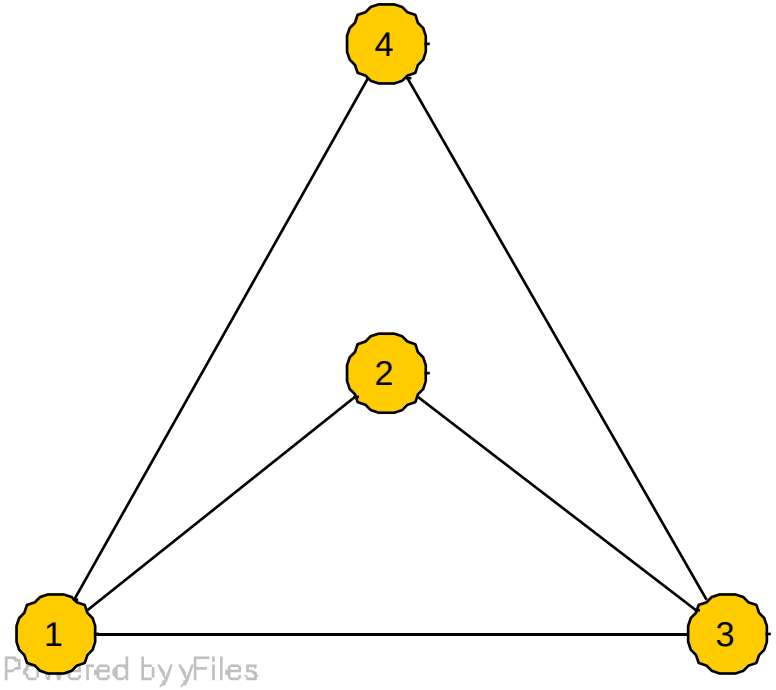
Graph theoretic - details

- Two types of lines:
  - *Arcs* – lines connecting an ordered pair of actors. An arc is a directional line
  - *Edges* – lines connecting an unordered pair of actors. An edge is a non-directional line
- If a graph contains arcs, then it is called a directed graph, or a *digraph*
- If lines have values (other than 1), then the graph is called a *valued* graph
- Useful for centrality, prestige, subgroups

Graph theoretic - caution

- Important – be careful when visually interpreting a graph. There are many equivalent ways of displaying exactly the same graph:





Sociometric

- A data set that contains the relations among a set of actors is called *sociometric*.
- Sociometric data can be stored (and displayed) in a sociomatrix
- A one-mode network can use a square sociomatrix where the rows and columns represent the actors
- Useful for computer algorithms, structural equivalence

Sociometric - continued

- The rows represent the sending actors, the columns represent the receiving actors (…*from* the row, *to* the column…)
- The cells of the matrix store the relational values
- X_{ij} = the value of the tie from n_i to n_j

Sociometric - example

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	0	1
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	0
Ross	0	0	0	0	-	1
Sarah	0	1	0	0	0	-

Algebraic/Set theoretic

- Uses *set theory* notation
  - $N = \{n_1, n_2, \ldots, n_g\} = \{\text{Allison}, \text{Drew}, \ldots, \text{Sarah}\}$
  - $L = \{ l_1, l_2, \ldots, l_L\} =$ $\{(\text{Allison}, \text{Drew}), (\text{Allison}, \text{Ross}), \ldots, (\text{Sarah}, \text{Drew})\}$
  - $G = \{N, L\}$
- Useful for multiple relations, developing more formal mathematical foundations of network methods (e.g., generalized blockmodels)

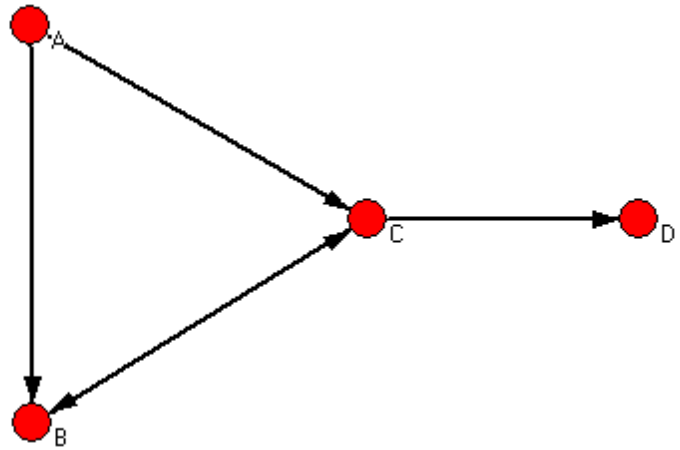
Edge list

- Can specify a complete network with a list of network edges
- Example:
  - Allison-Drew
  - Allison-Ross
  - Drew-Eliot
  - Drew-Sarah
  - Keith-Ross
  - Ross-Sarah
- Useful for efficient storage of very large networks

Network-data equivalency

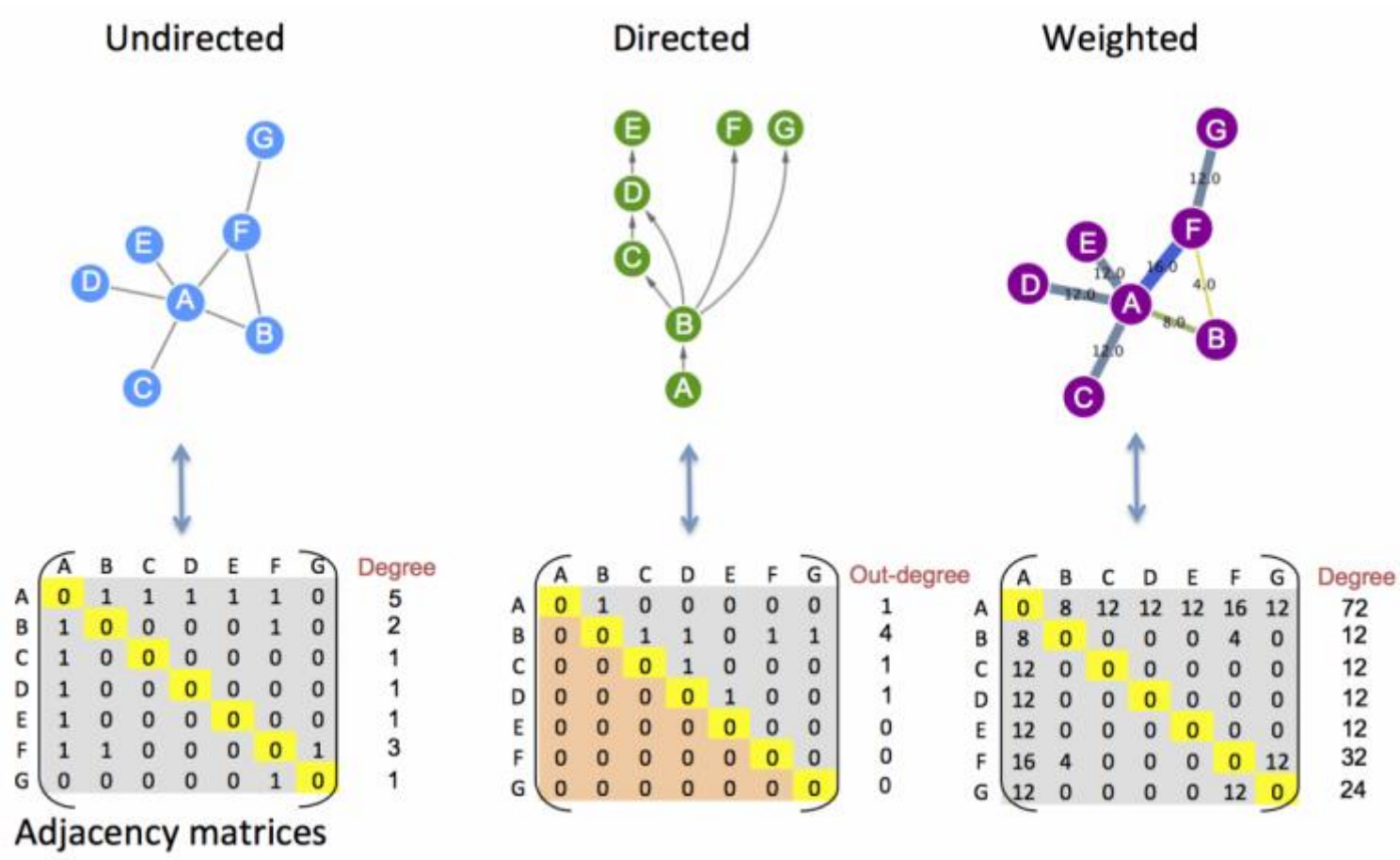
	A	B	C	D
A		1	1	0
B	0		1	0
C	0	1		1
D	0	0	0	

Sociomatrix



Graph

Graph types and sociomatrices



https://www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction

Building blocks for understanding network depiction and characteristics

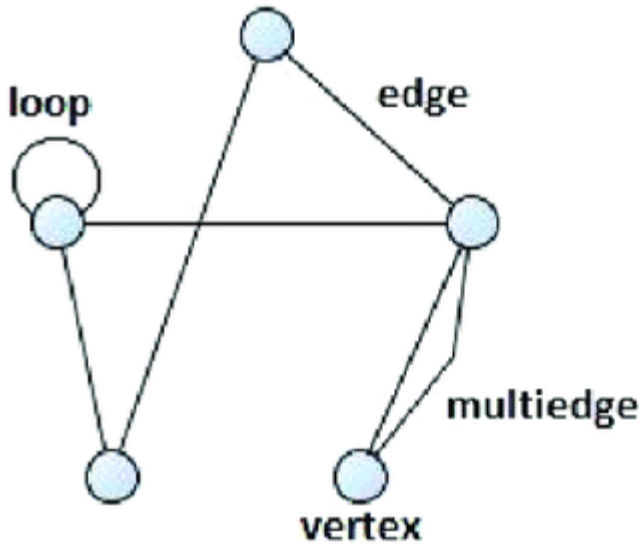
GRAPH THEORY

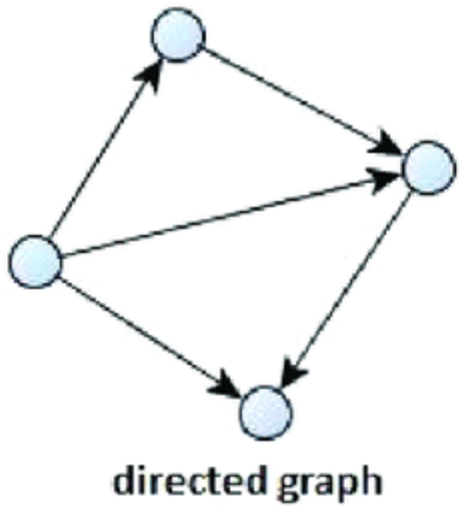
Non-directed graphs

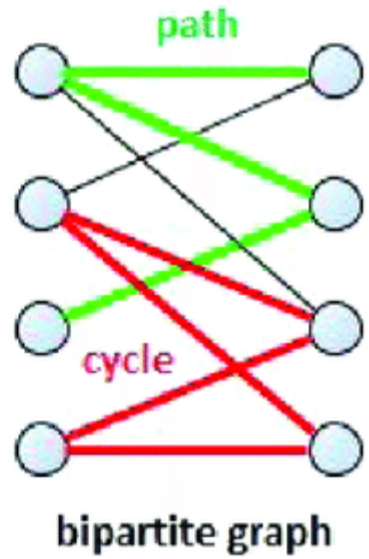
Edge	• line connecting two nodes (non-directed)
Loop	• line connecting a node to itself (often not allowed)
Simple graph	• no loops, and no more than one line between any pair of nodes
Adjacent	• adjacent nodes have a line connecting them
Incident	• a line connected to a node is incident to it
Complete	• all possible lines are present
Dyad	• two nodes, possibly connected
Triad	• three nodes, and their possible lines
Subgraph	• a subset of a graph, obtained by taking either a subset of the nodes or the lines of the graph

Non-directed graph connections

Degree	• number of lines incident to a node
Walk	• sequence of nodes and lines in which each node is incident with the lines preceding it and following it in the walk
Length	• length of a walk is the number of lines in it
Path	• a walk in which all lines and nodes are distinct
Reachable	• if a path exists between two nodes A and B, then A and B are reachable
Closed walk	• walk that begins and ends at the same node
Cycle	• closed walk where all lines and nodes (except the first) are distinct
Connected graph	• a connected graph has a path between every pair of nodes
Component	• a maximally connected subgraph







https://www.researchgate.net/publication/322253099_Small_Molecules-Big_Data

Non-directed graph connections

Component	• a maximally connected subgraph
Cutpoint	• a node that, if removed, would create additional components
Bridge	• a line that, if removed, would create additional components
Density	• in a graph (or subgraph), the proportion of observed ties to maximum possible number of ties

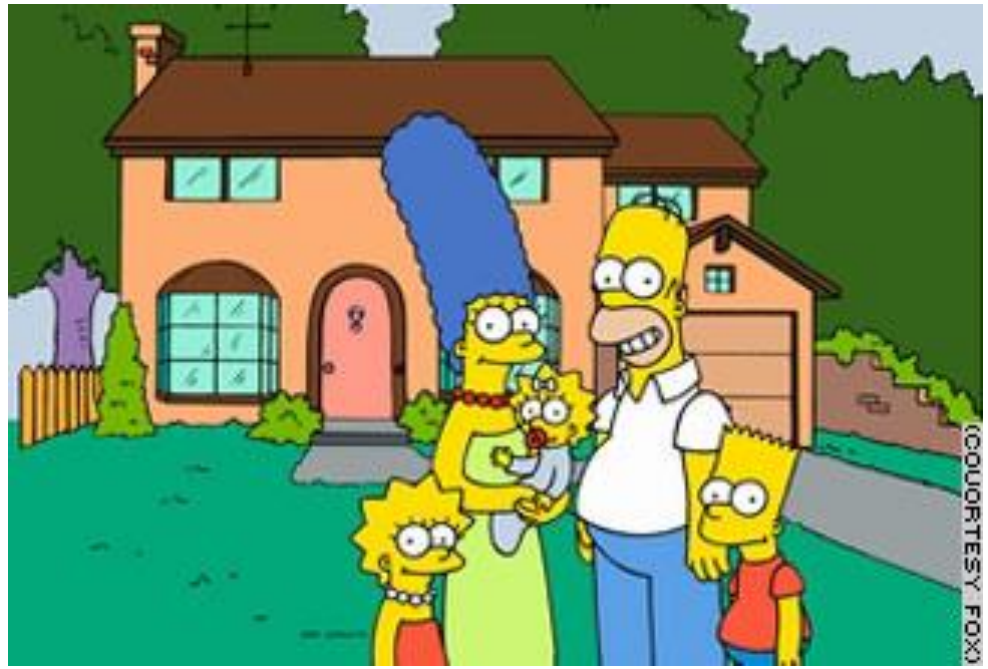
Non-directed graph distances

Geodesic	• shortest path between two nodes
Distance	• length of the geodesic
Diameter	• the diameter of a connected graph is the largest geodesic distance between any pair of nodes (i.e., the longest, shortest path!)

Directed graphs

Arc	• directed line connecting two nodes
Indegree	• number of arcs terminating at a particular node
Outdegree	• number of arcs originating at a particular node
Directed walk	• sequence of nodes and arcs, following the direction of the arcs
Directed path	• a directed walk with no repeating segments
Semipath, semiwalk	• removes the restriction of following the arcs in their normal directions
Reachable	• if there is a directed path from A to B, then B is reachable from A
Others	• digraphs lead to more complicated definitions of connectivity, geodesics, distance, diameter, etc. See W&F, pp. 132-134

Familiar network



Simpsons – size & density