Note on Graph and Network

Zepeng CHEN
The HK PolyU

Date: January 11, 2023

1 Notation of Graph

Definition 1.1 (Graph (Network))

A network (graph) (N, A, G) *consists of*

1. $N = \{1, ..., n\}$*: the node set (*n *nodes);*
2. A*: the arcs set (*m *arcs);*
3. G*:* $n \times n$ *adjacency matrix, where* $g_{ij} = 1$ *if* i *and* j *are connected, and* $g_{ij} = 0$ *otherwise.*

Note on Network Flow *In network flow problem, the 1st node we call it source node, and the last node we call it sink node.*

Note on *Social Network Analysis In SNA, we assume that there is no self-loops, i.e.,* $g_{ii} = 0$*. For example,*

1. *Empty network:* $g_{ij} = 0$ *for all* $i \neq j$*;*
2. *Complete network:* $g_{ij} = 1$ *for all* $i \neq j$*;*
3. *Directed network: there exists* g_{ij} *s.t.* $g_{ij} \neq g_{ji}$*;*
4. *Weighted network: there exists* g_{ij} *s.t.* $0 < g_{ij} < 1$*.*

Definition 1.2 (Subgraph)

Suppose $S \subseteq N$ *is the subset of nodes set, then a subgraph based on* S *is defined as*

$$G|_S = \{ij \mid g_{ij} = 1 \text{ and } i, j \in S\}.$$

Definition 1.3 (Degree)

Degree(i) is the total number of arc that are incident to node i.

$$indegree(i) + outdegree(i) = degree(i)$$

Note on *Network flow We can always find a node with indegree=0 (source node) and a node with outdegree=0 (sink node).*

Lemma 1.1

The sum of degrees is equal to twice the number of arcs, i.e. $\sum_{i \in N} d_i(G) = 2m$*.*

Definition 1.4 (Walk)

A walk in G is a G' consisting of nodes and arcs $i_1 - a_1 - i_2 - \ldots - i_{r-1} - a_{r-1} - i_r$ satisfying that $a_k = (i_k, i_{k+1}) \in A$ or $a_k = (i_{k+1}, i_k) \in A$.

Definition 1.5 (Path)

A path is a walk without repetition of nodes. A walk is more free and a path can not go back to points visited before.

Definition 1.6 (Loop)

A loop is an arc whose tail node is the same as its head node.

Definition 1.7 (Cycle)

A cycle is a path $i_1 - i_2 - \ldots - i_{r-1} - i_r$ together with arc (i_r, i_1) or (i_1, i_r).

Proposition 1.1 (Degree and cycle (Kanti, 2018))

A graph with all degrees greater than 2 contains a cycle.

Definition 1.8 (Geodesic)

A geodesic between i and j is a shortest path between i and j.

Definition 1.9 (Component)

A component of (N, G) is its maximal connected subnetwork: (N', G') s.t.

1. *subnetwork: $\emptyset \neq N' \subseteq N$ and $\mathrm{G}' \subseteq G$;*
2. *connectness: (N', G') is connected;*
3. *maximal: if $i \in N'$ and $ij \in G$, then $j \in N'$ and $ij \in G'$.*

Note on *The set of components of a network (N, g) given N is denoted $C(g)$, and the component containing a specific node i is denoted $C_i(g)$. Let $\Pi(N, g)$ denote the partition of N induced by the network (N, g). For example, p48. In other words, a network is connected iff it consists of a single component, i.e., $\Pi(N, g) = \{N\}$.*

Definition 1.10 (Bridge)

A link ij is a bridge in the network g if $g - ij$ has more components than g.

Definition 1.11 (Acyclic graph)

A graph is acyclic if it contains no directed cycle.

Definition 1.12 (Connectivity and Strong Connectivity)

A network is connected if every two nodes are connected by some path in the network. A connected graph is strongly connected if it contains at least one directed path from one node to every other node.

Definition 1.13 (Cut, s-t Cut)

A cut is a partition of the nodes in N into two parts: S and $\bar{S} = N - S$.

2 Special Network

Definition 2.1 (Tree Network)

A tree is a connected graph that contains no cycles.

Lemma 2.1

- *A tree on n nodes contains $n - 1$ arcs.*
- *A tree has at least two leaf nodes (degree=1).*
- *Every two nodes in a tree are connected by a unique path.*

Definition 2.2 (Forest)

A forest is a network such that each component is a tree. Thus any network that has no cycles is a forest.

Definition 2.3 (Cayley Tree)

Each node besides leaves has degree d.

Definition 2.4 (Rooted tree)

A rooted tree is a tree with a specifically designated node, called its root. Arcs in a rooted tree define predecessor-successor relationships.

Definition 2.5 (Direct-out tree)

A direct-out tree is a tree rooted in s, where the unique path in the tree from node s to every other node is a directed path.

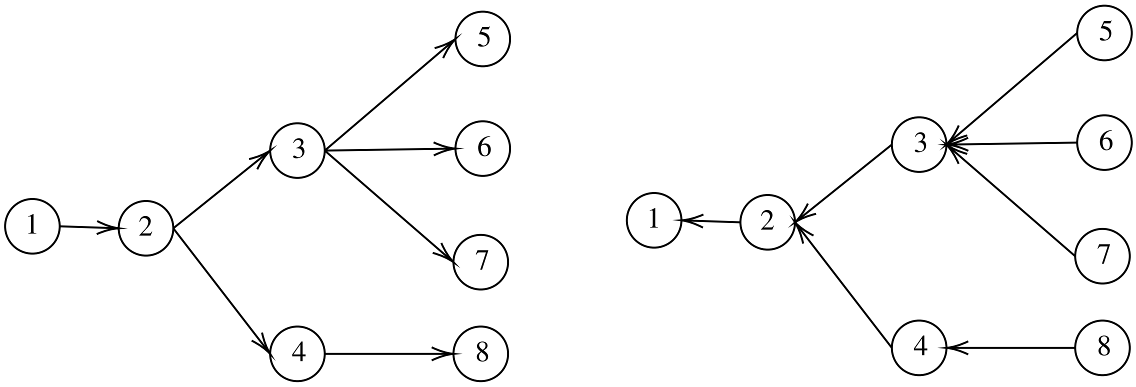

Figure 1: Instances of directed out-tree and directed in-tree.

Definition 2.6 (Spanning tree)

A tree T is a spanning tree of G if T is a spanning subgraph of G.

Theorem 2.1

Let $G = (N, \epsilon)$ be a connected undirected graph and let ϵ_0 be some subset of the set ϵ of arcs. Suppose that the arcs in ϵ_0 do not form any cycles. Then, the set ϵ_0 can be augmented to a set $\epsilon_1 \supset \epsilon_0$ so that (N, ϵ_1) is a spanning tree.[a]

[a] Bertsimas (1997). Introduction to linear optimization

Definition 2.7 (Fundamental cycle)

Addition of any non-tree arc to the spanning tree T creates exactly one cycle–fundamental cycle. Since a graph has $m - n + 1$ non-tree arcs, it has $m - n + 1$ fundamental cycles.

Definition 2.8 (Star Network)

(N, G) is a star network if it is a tree that has a "center" node i such that every link in the network involves node i.

Note on *Star network is a special case of tree.*

Definition 2.9 (Complete Network)

The complete network is a graph where all possible links are present, i.e. $g_{ij} = 1$ for all $i \neq j$.

Definition 2.10 (Neighborhood (friends))

1. *The neighborhood (friends) of node i is the set of nodes that i is linked: $N_i(G) = \{j : g_{ij} = 1\}$;*
2. *All the nodes that can be reached from i by paths of length no more than k is the k-neighborhood of i, i.e.*

$$N_i^k(g) = N_i(g) \cup \left(\bigcup_{j \in N_i(g)} N_j^{k-1}(g) \right);$$

3. *Given a set of nodes S, the neighborhood of S is the union of the neighborhoods of its members, i.e.*

$$N_S(g) = \bigcup_{i \in S} N_i(g) = \{j : \exists i \in S, g_{ij} = 1\}.$$

Definition 2.11 (Circle Network)

(N, G) is a circle network if it has a "single" cycle and each node in the network has exactly two neighbors.

3 Statistics of Network

Definition 3.1 (Degree distribution)

The degree distribution of a network is a description of the relative frequencies of nodes that have different degrees.

Definition 3.2 (Distance, diameter, Average path length)

1. *The distance between two nodes is the length of the shortest path or geodesic between them.*
2. *The diameter of a network is the largest distance between any two nodes in the network.*
3. *Average path length is the average taken over geodesics.*

Definition 3.3 (Cliquishness)

$S \subseteq N$ is clique if

1. *$G|_S$ is a complete network;*
2. *for any $i \in N \backslash S$, $G|_{S \cup \{i\}}$ is not complete.*

Note on *A clique is a maximal completely connected subnetwork.*

Definition 3.4 (Clustering measure)

If i is a friend of j and k, CL measures how likely is it on average j and k are friend:

$$CL(G) = \frac{\sum_i \# \{jk \in G \mid j \neq k, j, k \in N_i(G)\}}{\sum_i \# \{jk \mid j \neq k, j, k \in N_i(G)\}}$$

Note on *The most common way to measure some aspect of cliqueshness.*

Definition 3.5 (Degree centrality)

Degree centrality of i is $\frac{1}{n-1} \sum_{j \in N} g_{ij}$.

Note on *Clustering measure is in macro level, and degree centrality is in micro level. It tells us how well a node is connected, in terms of direct connections.*

Definition 3.6 (Decay centrality)

Decay centrality of i is $\sum_{j \neq i} \delta^{l(i,j)}$, where $\delta \in (0,1)$ is the decay parameter, and $l(i,j)$ is the number of links in the shortest path between i and j.

Definition 3.7 (Betweenness centrality)

Betweenness centrality of i is $\sum_{k \neq j : i \notin \{k,j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2}$, where $p_i(kj)$ is the number of geodesics between k and j that involves i, and $P(kj)$ is the number of geodesics between k and j.

4 Network Representation

Definition 4.1 (Node-Arc incidence matrix)

- *Arc (i, j) has +1 in row i and -1 in row j.*
- *Storage efficiency $\frac{2m}{nm}$ and becomes inefficient when n is large.*
- *Sum of all rows is zero, it means these rows are not independent.*

Let us now focus on the ith row of A (node-arc incidence matrix), denoted by a_i'. Thus

$$\mathbf{a}_i'\mathbf{f} = \sum_{j\in O(i)} f_{ij} - \sum_{j\in I(i)} f_{ji} = b_i \quad \text{or} \quad Af = b$$

Definition 4.2 (Node-Node adjacency matrix)

x

Proposition 4.1 (Power of Adjacency matrix)

Let A denotes the node-node adjacency matrix of a network G. Then the ijth entry of the matrix $A^k, k = 1, ..., n$ is the number of walks of length k from node i to node j.[a]

[a] Finding path-lengths by the power of Adjacency matrix of an undirected graph

Proof By induction, ■

Proposition 4.2 (Adjacency matrix and strongly connected[1])

Let A denotes the node-node adjacency matrix of a network G. Then G is strongly connected iff the matrix $A + A^2 + ... + A^n$ has no zero entry.

Definition 4.3 (Forward Star Representation)

Definition 4.4 (Backward Star Representation)

5 Hall's theorem and Bipartite Graphs

6 Set coverings and Independent Sets

Definition 6.1 (Independent set)

An independent set of nodes $A \subseteq N$ is a set such that if $i \in A$, $j \in A$, and $i \neq j$, then $ij \notin G$.

Definition 6.2 (Maximal Independent set)

An independent set of nodes A is maximal if it is not a proper subset of any other independent set of nodes.

Lemma 6.1

Consider a network (N,g) and a network (N,g') such that $g \subset g'$*. Any independent set* A *of* g' *is an independent set of* g*, but if* $g' \neq g$ *then there exist (maximal) independent sets of* g *that are not (maximal) independent sets in* g'*.*

7 Colorings

8 Eulerian Tours and Hamiltonian Cycles

Definition 8.1 (Eulerian Cycles)

x

Theorem 8.1 (Eulerian Cycles theorem)

A graph (N, G) *has an Eulerian cycle iff each node must have an even degree.*

Proof

■

Definition 8.2 (Hamiltonian cycle)

Lemma 8.1 (Hamiltonian cycle theorem)

Proof

■

Bibliography

Kanti, John (Dec. 2018). *Proof That If All Vertices Have Degree at Least Two Then G Contains a Cycle*. Forum Post.