Network Theory

April 23, 2008

Contents

Chapter 1

Introduction to Networks

There are many different types of networks. Such as networks mapping the internet (for example, see `http://www.opte.org/maps/`), networks mapping the food chain, networks mapping the collaboration between scientists, business people (for example, see: `http://theyrule.net`), and many more.

1.1 Definition

Definition A network or graph is defined as a collection of n nodes connected by m edges. A network can be directed, meaning the edges point in one direction, or undirected, meaning the edges go in both directions. The edges can join more than two vertices together. Such graphs are called hypergraphs.The edges can be weighted, contain self loops, and have different properties within the edges or nodes.

1.2 Types of Networks

- **Social networks**: friendship networks
- **Organizational**: business partnerships
- **Biological networks**: protein networks
- **Technological networks**: electrical grid, telephone calls
- **Naturally Occurring networks**: friendship networks, food chain.

1.3 Research Problems and Areas

There are many research problems that involve network theory such as statistical properties, finding information in a network, epidemics, passing a message, predicting robustness of a network, e.g., what nodes are the most vulnerable; what Scale-free properties (e.g., why not Poisson, uniform, or normal), and Small world properties. Most issues address huge graphs where $n > 10^9$.

Network theory is found in Computer Science through graph theory, artificial intelligence, natural language processing, computer networks, and information retrieval. It is also found in many other fields such as in statistical mechanics in physics, probabilities in statistics, social networks in sociology, linguistic networks in linguistics, electrical engineering, ecology, and business.

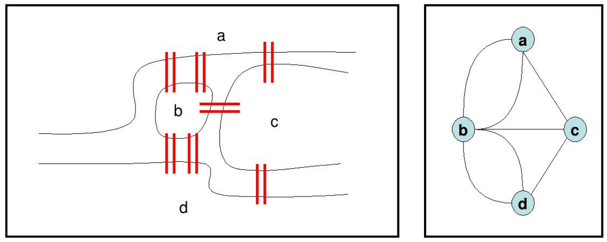


Figure 1.1: Seven Bridges of Konigsberg and the corresponding graph

1.4 Euler's Problem

Euler's Problem is an example of a graph problem. It was first analyzed while Leonhard Euler was solving the famous Seven Bridges of Konigsberg problem in 1736 (See figure 1.1). The problem was to find a way to cross all bridges without crossing over any bridge twice. An Euler tour is a graph which visits each edge exactly once and starts and ends at the same vertex. It is an NP-Complete problem. More information can be found at `http://en.wikipedia.org/wiki/Eulerian_path`.

1.5 Network Terminology

A graph (or network) G = (V,E) is a data structure composed of a set of vertices (V) (or nodes, site, actor) and a set of edges (E) (or link, bond, tie) connecting the vertices in V.

An edge e = (u,v) is a pair of connected vertices.

If (v_0, v_1) is an edge in an undirected graph, v_0 and v_1 are **adjacent**. The edge (v_0, v_1) is **incident** on vertices v_0 and v_1.

If (v_0, v_1) is an edge in a directed graph, v_0 is **adjacent to** v_1, and v_1 is **adjacent from** v_0. The edge (v_0, v_1) is **incident** on v_0 and v_1.

The **degree** of a vertex is the number of edges incident to that vertex. In directed graphs, the **in-degree** of a vertex v is the number of edges that have v as the head and the **out-degree** of a vertex v is the number of edges that have v as the tail.

A **path** is a sequence of vertices $v_1, v_2, \ldots, v_k$ such that consecutive vertices $v_i, andv_{i+1}$ are adjacent. A simple path is one with no repeated vertices and a **cycle** is a simple path except the last vertex is the same as the first vertex.

A **connected graph** is a graph where any two vertices are connected by some path.

A **subgraph** is a subset of vertices and edges forming a graph. A **connected component** is a maximal connected subgraph.

An **undirected graph** is one in which the pair of vertices in a edge is unordered, $(v_0, v_1) = (v_1, v_0)$ and a **directed graph** is one in which each edge is a directed pair of vertices, $(v_0, v_1) \neq (v_1, v_0)$

Edges can have weights associated with them.

An **assortative network** is a network where nodes of high degree connect to nodes of high degree. A **disassortative network** is a network where of high degree connect to nodes of low degree.

1.6 Graph Representations

1.6.1 Adjacency Matrix

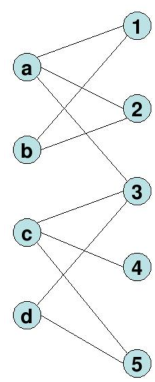
Let G=(V,E) be a graph with n vertices. The adjacency matrix of G is a two-dimensional n x n array, say adj_mat. If the edge (v_i, v_j) is in E(G), adj_mat[i][j]=1. If there is no such edge in E(G), adj_mat[i][j]=0. The adjacency matrix for an undirected graph is symmetric; the adjacency matrix for a digraph need not be symmetric. An adjacency matrix has fast access to adjacent nodes, and fast computation of in- and out-degree, but it is memory inefficient. A graph with multiple components will contain separate groupings of 1's.

1.6.2 Adjacency List

Each row in an adjacency matrix is represented as a list. An undirected graph with n vertices and e edges => n head nodes and $2e$ list nodes. An adjacency list is a more compact representation and therefore memory efficient. However, sometimes it has less efficient computation of in- or out- degrees.

1.7 Bipartite Graph

A bipartite graph is a graph whose vertices can be divided into two disjoint sets s_1 and s_2 such that each edge connects a vertex from s_1 to a vertex from s_2. Figure 1.2 is an example of a bipartite graph and matrices representing the number of vertices in the other set that join the two vertices in the same set. For example, vertex b is connected to vertex a twice; through vertex 2 and 1.



	a	b	c	d
a	-	2	1	1
b	2	-	0	0
c	1	0	-	2
d	1	0	2	-

	1	2	3	4	5
1	-	2	1	0	0
2	2	-	1	0	0
3	1	1	-	1	2
4	0	0	1	-	1
5	0	0	2	1	-

Figure 1.2: Bipartite Graph

1.8 Graph Planarity

The Kuratowski theorem states that a graph G is planar iff it doesn't include a homeomorphism of K_5 or $K_{3,3}$ (See Figure 1.3). There is an algorithm that runs linear in N called the Boyer-Myrvold planarity algorithm. See the planarity game at `http://www.planarity.net`.

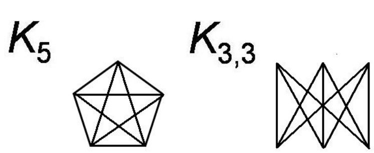


Figure 1.3: A graph is planar iff it doesn't include these homeomorphisms

1.9 Readings

1. Sections I, II, III of [1] (pages 48-59 or 1-14 of PDF)

2. Sections I, II, IV of [5] (pages 1-9 and 20-26)

Chapter 2

Real Networks

Lecturer: Dragomir Radev
Scribe: Sara Stolbach

2.1 Network Examples

- Interdisciplinary collaborations on papers mentioned in [3]
- Yeast proteins mentions in [4]
- High School Dating in [2] and viewable at http://www.soc.duke.edu/~jmoody77/NetMovies/rom_flip.htm
- Free word association in [6]
- New York Subway Map
- Citation Networks
- Social network analysis software, http://www.orgnet.com/index.html such as Political Polarization, http://www.orgnet.com/divided.html, and Email Communication by Valdis Krebs, http://www.orgnet.com/email.html
- Multiple network graphics, http://www.visualcomplexity.com/vc/ such as Presidential Watch 2008 at http://www.visualcomplexity.com/vc/project_details.cfm?id=542\&index=542\&domain, a Working Brain Model at http://www.visualcomplexity.com/vc/project_details.cfm?id=530\&index=530\&domain=, A social network visualization at http://www.visualcomplexity.com/vc/project_details.cfm?id=524\&index=524\&domain=, and Call and Response at http://www.visualcomplexity.com/vc/project_details.cfm?id=535\&index=535\&domain=
- Additional network graphics, http://www.soc.duke.edu/~jmoody77/NetMovies/index.htm
- Flight Patterns, http://www.youtube.com/watch?v=dPv8psZsvIU
- Additional network graphics, http://vw.indiana.edu/07netsci/entries/
- Idea mapping, http://images.businessweek.com/ss/07/11/1115_in_network/index_01.htm
- Citation pattern between the top political blogs, http://www.hpl.hp.com/research/idl/demos/politicalblogdemo.html
- Growing a network with random and preferential attachment, http://projects.si.umich.edu/netlearn/NetLogo/PrefAndRandAttach.html

- Diffusion of a small world model, http://projects.si.umich.edu/netlearn/NetLogo/SmallWoldDiffusion.html
- Search in a 2D world, http://projects.si.umich.edu/netlearn/GUESS/2Dsearch_r1.html
- The betweenness clustering algorithm, http://projects.si.umich.edu/netlearn/GUESS/betweennessclust.html

2.2 Large Network

> "Initial production of the A380 was plagued by delays attributed to the 530 km (330 miles) of wiring in each aircraft. Airbus cited as underlying causes the complexity of the cabin wiring (100,000 wires and 40,300 connectors), its concurrent design and production, the use of two incompatible versions of the CATIA computer-aided design software, the high degree of customization for each airline, and failures of configuration management and change control. Deliveries would be pushed back by nearly two years." http://en.wikipedia.org/wiki/Airbus_A380.

2.3 Erdos Number

The Erdos number is the number of hops it takes to get to a paper published by Paul Erdos. 511 people have E=1 and 8,000 have E=2. For example, the path from Paul Erdos to Dragomir Radev is:

> Erdos - Grunbaum (1972) → Grunbaum - Sloane (1974) → Sloane - Aho (1973) → Aho - Radev (1998).

Therefore, Dragomir Radev has an Erdos number of E=4.

2.4 Kevin Bacon

The six degrees of kevin bacon game is the game of how many hops it takes for any actor to get to Kevin Bacon , http://oracleofbacon.org/, http://www.cinfn.com/, http://www.thekevinbacongame.com/

2.5 Readings

1. Sections I, II, III of [1] (pages 48-59 or 1-14 of PDF)
2. Sections I, II, IV of [5] (pages 1-9 and 20-26)

References

[1] Réka Albert and Albert-László Barabási. “Statistical Mechanics of Complex Networks”. In: *RMP* 74.1 (2002). Pp. 47–98.

[2] Peter S. Bearman, James Moody, and Katherine Stovel. “Chains of affection: The structure of adolescent romantic and sexual networks”. In: *American Journal of Sociology* 110.3 (2004). Pp. 44–91.

[3] M. Girvan and M. E. J. Newman. “Community structure in social and biological networks”. In: *PNAS* 99.12 (2002). Pp. 7821–7826.

[4] Sergei Maslov and Kim Sneppen. “Specificity and Stability in Topology of Protein Networks”. In: *Science* 296.5569 (2002). Pp. 910–913.

[5] Mark E. J. Newman. “The Structure and Function of Complex Networks”. In: *SIAM Review* 45.2 (2003). Pp. 167–256.

[6] Mark Steyvers and Joshua B. Tenenbaum. “The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth”. In: *Cognitive Science* 29.1 (2005). Pp. 41–78.