Graphs

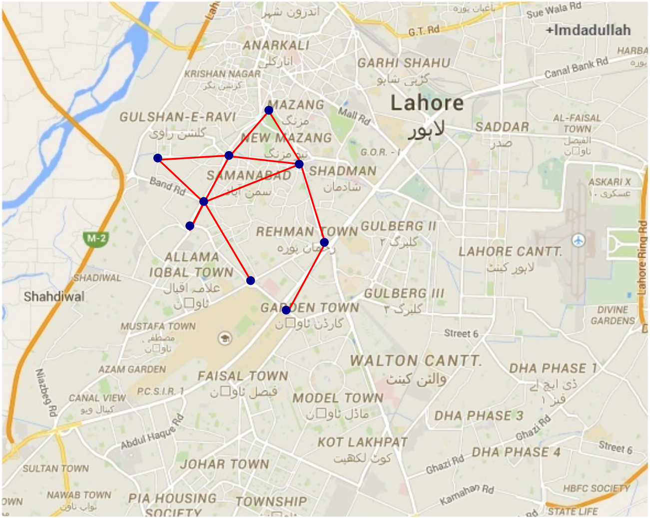
- Graphs are everywhere
- Graph Types and Terminology: Handshaking Lemma
- Graph Representation, Complement, Transpose, Subgraph
- Walks, Paths and Cycles
- (Strongly) Connected Graphs, k-Connected Graphs
- Graphs Applications: BFS, DFS, Eulerian Graphs
- Advanced Applications of Graphs Algorithms and Analysis

Imdad ullah Khan

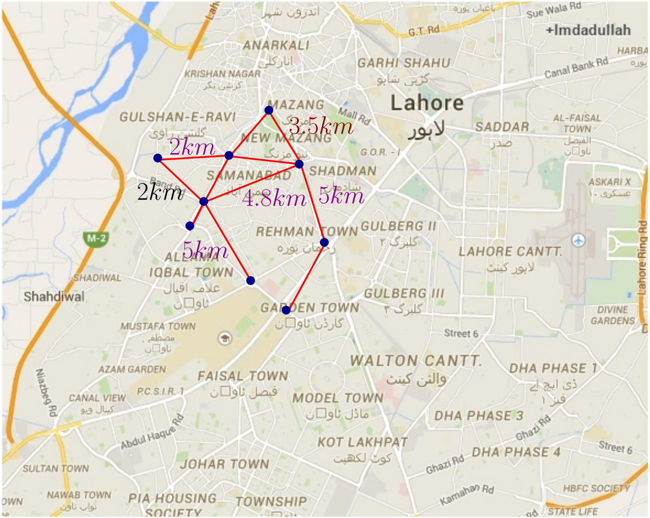
Graphs Applications: Shortest Path

Graphs Applications: Shortest Path

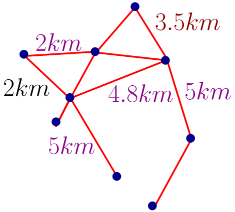
Graphs Applications: Shortest Path



Graphs Applications: Shortest Path



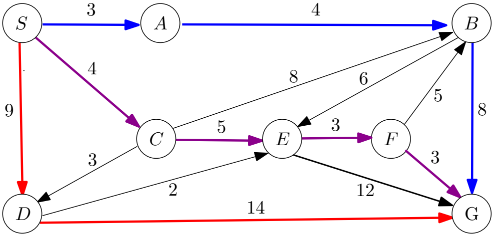
Graphs Applications: Shortest Path



Weight of Paths

Weight or length of a path $p = v_0, v_1, \ldots, v_k$ in weighted graphs is sum of the weights of its edges

$$C(p) \; = \; \sum_{i=1}^{k} w(v_{i-1}, v_i)$$



Three $S - G$ paths

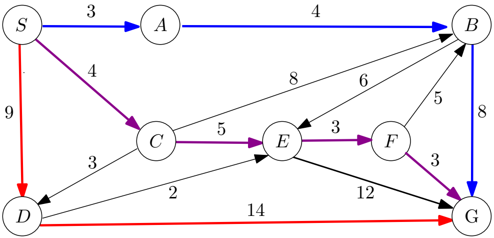
$\mathbf{C(p_1) = 3 + 4 + 8}$

$\mathbf{C(p_2) = 4 + 5 + 3 + 3}$

$\mathbf{C(p_3) = 9 + 14}$

Unweighted graphs are weighted graphs with unit edge weights

Shortest Paths



Three $S - G$ paths

$\mathbf{C(p_1) = 3 + 4 + 8}$

$\mathbf{C(p_2) = 4 + 5 + 3 + 3}$

$\mathbf{C(p_3) = 9 + 14}$

Shortest path from s to t is a path of smallest weight

Distance from s to t, $\mathbf{d(s, t)}$: weight of the shortest $s - t$ path

▷ There can be multiple shortest $s - t$ paths

Graphs Applications: Shortest Path Problems

Shortest Path Problems

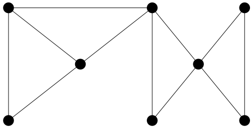
Given a weighted graph, $G = (V, E, w)$

1. Find a shortest path from a vertex s to another vertex t
2. Find shortest paths from a vertex s to all other vertices
3. Find shortest paths from each vertex to every other vertex

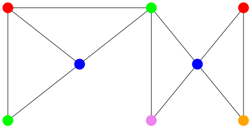
Use Dijkstra Algorithm, Bellman-Ford algorithm, Floyd-Warshal Algorithm

Graph Coloring

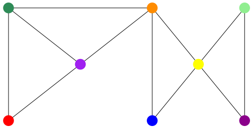
A graph (vertex) coloring is to assign a color to each vertex such that no two adjacent vertices get the same color



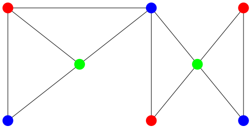
A graph G on 8 vertices



A coloring with 6 colors



A coloring with 8 colors

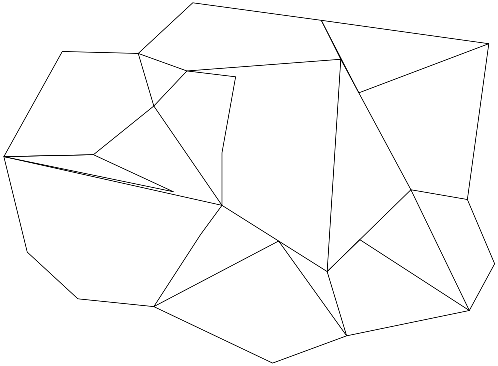


A coloring with (optimal) 3 colors

Want to use the minimum number of colors

- ▷ Minimum colors needed to color G is chromatic number of G, $\chi(G)$

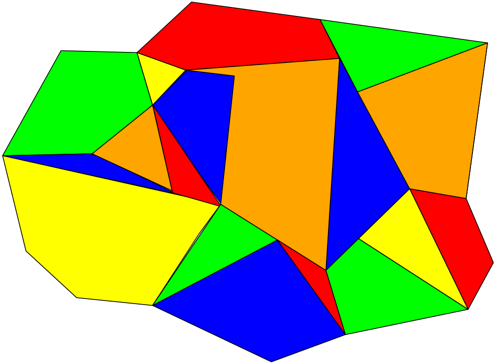
Graphs Applications: Coloring



Color regions of map

No neighboring regions can have the same color

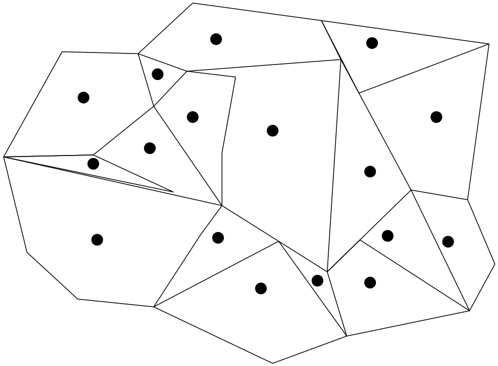
Graphs Applications: Coloring



Color regions of map

No neighboring regions can have the same color

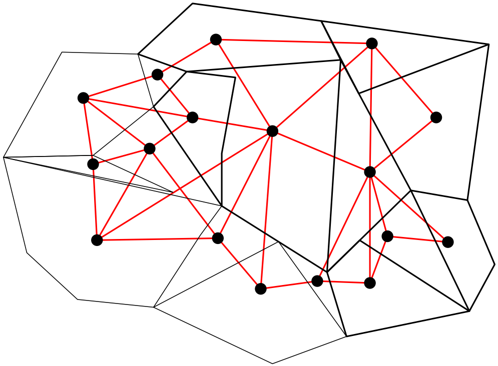
Graphs Applications: Coloring



Color regions of map

No neighboring regions can have the same color

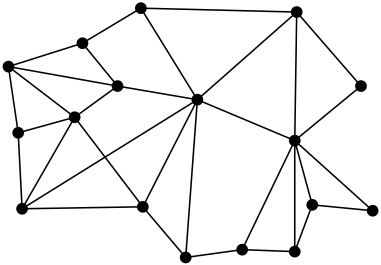
Graphs Applications: Coloring



Color regions of map

No neighboring regions can have the same color

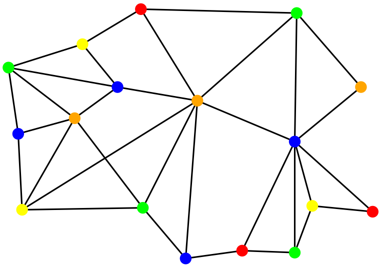
Graphs Applications: Coloring



Color vertices of G

No neighboring vertices can have the same color

Graphs Applications: Coloring



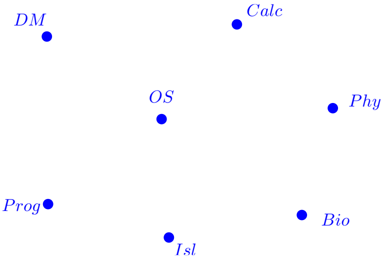
Color vertices of G

No neighboring vertices can have the same color

Graphs Applications: Coloring

Final Exam Scheduling

- No student should have two exams at a given time-slot
- How many time-slots are needed?

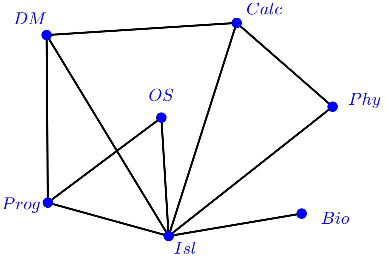


Make each course a vertex

Graphs Applications: Coloring

Final Exam Scheduling

- No student should have two exams at a given time-slot
- How many time-slots are needed?



Make each course a vertex

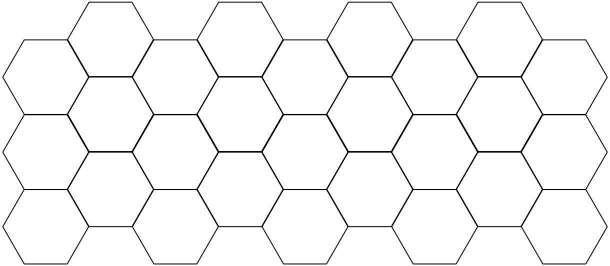
if two courses have a common student make an edge

Number of time-slots needed $= \chi(G)$

Graphs Applications: Coloring

Map Coloring and GSM Network

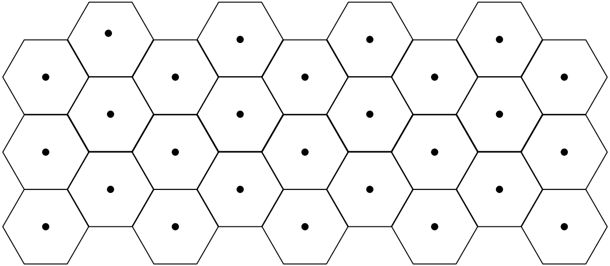
- In cellular networks (GSM) coverage area is divided into a hexagonal grid
- Each cell (a hexagon) is served by an antenna
- Each cell uses a frequency band (one of 850, 900, 1800, 1900 MHz)
- Frequency of a cell must be different from adjacent cells (hexagons sharing a line segment)
- Four color vertices of the dual graph of the hexagonal grid



Graphs Applications: Coloring

Map Coloring and GSM Network

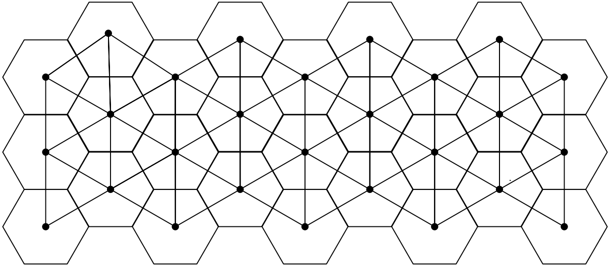
- In cellular networks (GSM) coverage area is divided into a hexagonal grid
- Each cell (a hexagon) is served by an antenna
- Each cell uses a frequency band (one of 850, 900, 1800, 1900 MHz)
- Frequency of a cell must be different from adjacent cells (hexagons sharing a line segment)
- Four color vertices of the dual graph of the hexagonal grid



Graphs Applications: Coloring

Map Coloring and GSM Network

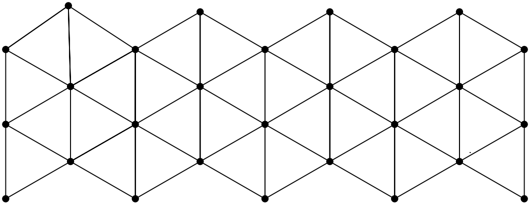
- In cellular networks (GSM) coverage area is divided into a hexagonal grid
- Each cell (a hexagon) is served by an antenna
- Each cell uses a frequency band (one of 850, 900, 1800, 1900 MHz)
- Frequency of a cell must be different from adjacent cells (hexagons sharing a line segment)
- Four color vertices of the dual graph of the hexagonal grid



Graphs Applications: Coloring

Map Coloring and GSM Network

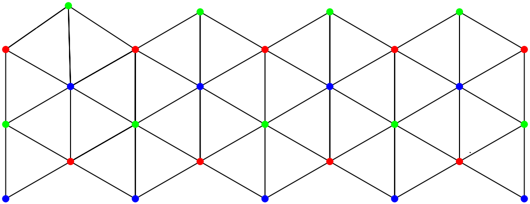
- In cellular networks (GSM) coverage area is divided into a hexagonal grid
- Each cell (a hexagon) is served by an antenna
- Each cell uses a frequency band (one of 850, 900, 1800, 1900 MHz)
- Frequency of a cell must be different from adjacent cells (hexagons sharing a line segment)
- Four color vertices of the dual graph of the hexagonal grid



Graphs Applications: Coloring

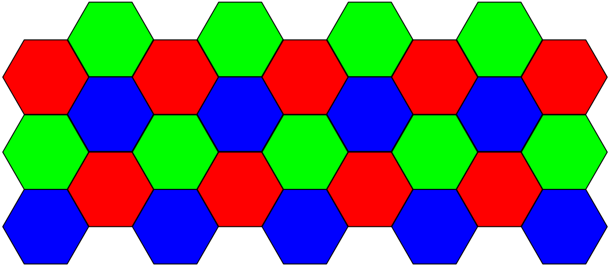
Map Coloring and GSM Network

- In cellular networks (GSM) coverage area is divided into a hexagonal grid
- Each cell (a hexagon) is served by an antenna
- Each cell uses a frequency band (one of 850, 900, 1800, 1900 MHz)
- Frequency of a cell must be different from adjacent cells (hexagons sharing a line segment)
- Four color vertices of the dual graph of the hexagonal grid



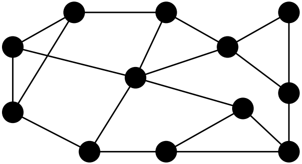
Map Coloring and GSM Network

- In cellular networks (GSM) coverage area is divided into a hexagonal grid
- Each cell (a hexagon) is served by an antenna
- Each cell uses a frequency band (one of 850, 900, 1800, 1900 MHz)
- Frequency of a cell must be different from adjacent cells (hexagons sharing a line segment)
- Four color vertices of the dual graph of the hexagonal grid

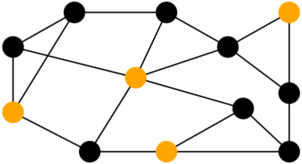


Independent Set in Graph

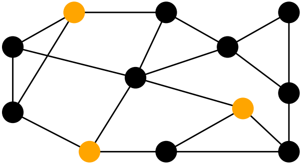
An independent set in G is subset of vertices no two of which are adjacent



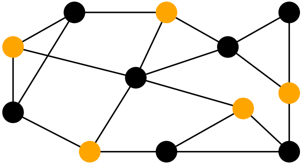
A graph on 12 vertices



An independent set of size 4



An independent set of size 3



An independent set of size 5 (max)

We want to find a large (the largest) independent set

Graphs Applications: Independent Set

Sites Selection Problem

- Suppose n potential sites are identified for opening up restaurants
- Some pairs of places shouldn't have the franchises at both of them (too close to each other, competitions, or operational constraints)
- Selecting k places for restaurants is selecting an independent set in a graph with vertices as sites and edges as pairwise competitions

The SNP (Single Nucleotide Polymorphism) Assembly Problem

- In computational biology (biochemistry) given a set of sequences we want to resolve inter-sequential conflicts by excluding some sequences
- Conflict between two sequences is due to their biochemical properties
- The goal is to select a large number of conflict free sequence
- This means in a graph with vertices representing sequences and edges representing conflicts, we want to find a large independent set

Graphs Applications: Independent Set

Diversifying Investment Portfolio

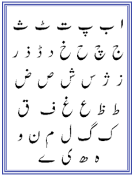
- Different stocks in a market
- $P_i(t)$ is price for stock i at time t
- $R_i(t) \; = \; \log \dfrac{P_i(t)}{P_i(t-1)}$, return or trading volume of stock i at time t
- Each stock is a node and two stocks have edges if correlation of their returns is $\geq \theta$ for threshold $-1 \leq \theta \leq 1$
- θ is set depending on potential risk (degree of diversification)
- Two adjacent vertices in $G_{\theta=.9}$ represent high risk investment pair

Set $\theta < -0.5$: an independent set in G_θ represents a portfolio with "small" risk (diverse set of investments)

Graphs Applications: Independent Set

Shannon Capacity of a graph

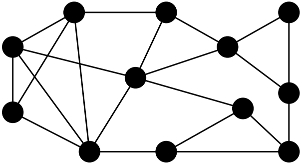
- Sending a message from an alphabet through a noisy channel



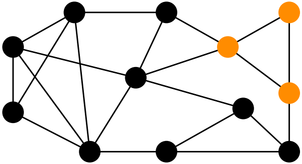
- Because of noise some characters can be confused
- How many 1 length strings can be sent without confusion?
- Make each letter a node and an edges iff the corresponding letters can be confused (depends on the SNR of channel)
- Max number of messages is the size of max independent set
- How many k-length strings can be sent on this channel?
- Size of max independent set in G^k (strong product of graphs)

Cliques in Graphs

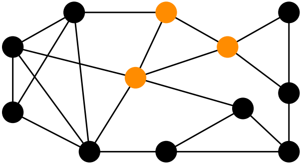
A clique in G is a subset of vertices every two of which are adjacent



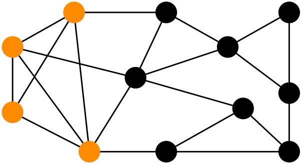
A graph on 12 vertices



A clique of size 3



A clique of size 3



A clique of size 4 (max)

We want to find a large (the largest) clique

Graphs Applications: Cliques

Cliques in Market Graphs

- Different stocks in a market
- $P_i(t)$ is price for stock i at time t
- $R_i(t) \;=\; \log \dfrac{P_i(t)}{P_i(t-1)}$, return or trading volume of stock i at time t
- Each stock is a node and two stocks have edges if correlation of their returns is $\geq \theta$ for threshold $-1 \leq \theta \leq 1$
- θ is set depending on potential risk (degree of diversification)
- Two adjacent vertices in $G_{\theta=.9}$ represent high risk investment pair

Set $\theta > 0.5$: a clique in G_θ represents a portfolio with "large" risk (diverse set of investments)

Can be of interest to a regulatory body - can be some kind of collusion

Graphs Applications: Cliques

Organized Tax Fraud Detection by IRS

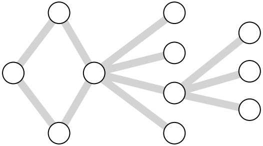
- Clustering similar objects is widely used in many applications
- Ideal clusters are cliques in a graph (community, highest internal degrees, lowest internal distances, largest internal densities etc.)
- Groups of phony tax returns are submitted to get undeserved returns
- IRS constructed graph, where each form is a vertex
- Edges between two vertices means similarity between the two forms is above some threshold
- A large clique in this graph points to a potential fraud

Location Covering Using Clique Partition

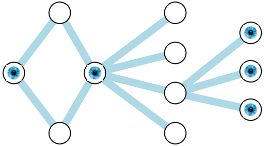
Protein Docking Problem

Vertex Cover

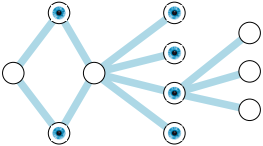
An vertex cover in a graph is subset C of vertices such that each edge has at least one endpoint in C



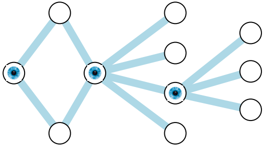
A graph on 11 vertices



A vertex cover of size 5



A vertex cover of size 6



A vertex cover of size 3

We look for a small(est) vertex cover

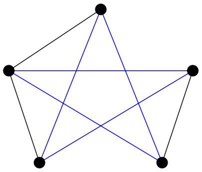
Graphs Applications: Vertex Cover

Network Security: Rout Based Filtering

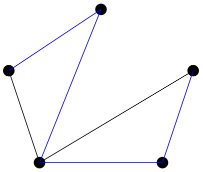
- Identify a small set of routers/AS
- So as all packets can be monitored at those routers/switches (check if the source/destination addresses is valid given the routing table and network topology)
- Route-based distributed packet filtering
- Prevents distributed denial of service attacks

(Directed) Hamiltonian Cycle and Path

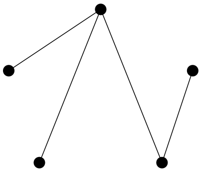
A Hamiltonian Cycle (path) in a graph is cycle (path) containing every vertex



Hamiltonian cycle in G



No Hamiltonian cycle in G
Hamiltonian path in blue



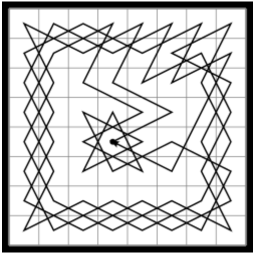
No Hamiltonian path in G
So no hamiltonian cycle

Unlike Euler circuit there is no if and only if type statement for existence of Hamiltonian cycle

Hamiltonian Cycle and Path Applications

Re-entrant Knight's Tours

- Is there a sequence of moves that takes the knight to each square on an 8×8 chessboard exactly once, returning to the original square
- For 8×8 AbuBakr Muhammad b. Yahya al-Suli discovered one in 9th century
- For $n \times n$ chessboard define a vertex for each position and connect vertex v_{ij} to vertex v_{kl} if there is a legal move between the i, jth position to the k, lth position on the board
- Find a Hamiltonian cycle in the graph



Graphs Applications: Hamiltonian Cycle

Genome Mapping

- Combine many tiny fragments of genetic codes (called "reads"), into one genomic sequence
- Consider each read a node in a graph
- Overlap (end of one read matches the start of another) is an edge
- A Hamiltonian cycle in this graph is a mapping of genomes

Route for School Bus

- Houses considered nodes and streets as edges
- Find a Hamiltonian cycle for a route visiting each student's house exactly once to save fuel and time

Longest Path Problem

Given a weighted graph, $G = (V, E, w)$

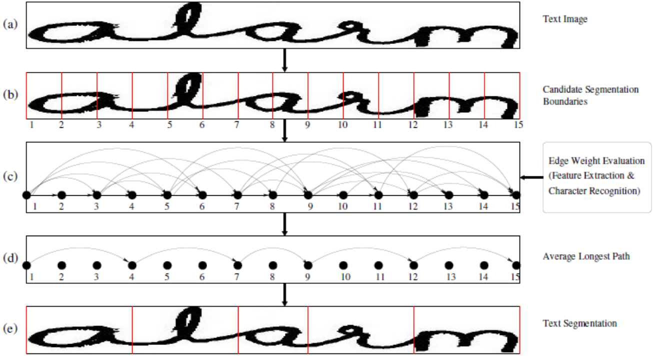
- Find a longest path from a vertex s to another vertex t.

The longest $s - t$ path P is a path from s to t with maximum total weight

Longest Path Problem: Application

Character Segmentation for Optical Character Recognition:

- First step in any OCR system is character segmentation
- Isolate individual characters in hand-written text
- Input to character recognition system
- Salvi et.al. (2013) proposed algorithm based on average longest paths



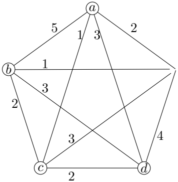
Longest Path Problem: Application

Static timing analysis (STA):

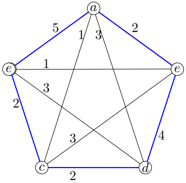
- Widely used method in circuit design and embedded systems
- STA: a simulation method of computing the expected timing of a digital circuit without requiring to simulate the full circuit
- Longest path identifies critical paths in an IC or VLSI system
- STA is performed only on these critical paths.

Traveling Salesman Problem (TSP)

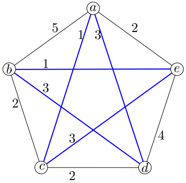
Given a complete graph G on n vertices with edge weights, a TSP tour is a Hamiltonian cycle in G



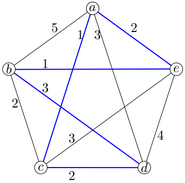
K_5 with edge weights



A TSP tour of length 15



A TSP tour of length 11

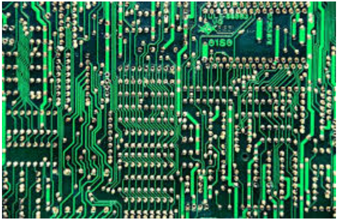


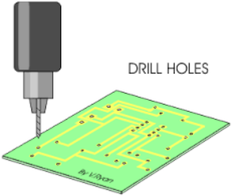
A TSP tour of length 9

Traveling Salesman Problem $\mathrm{TSP}(G)$: Find a minimum cost Hamiltonian cycle in G

Graphs Applications: TSP

- **Transportation:** A salesman wants to visit all cities with minimum cost
- Optimize the tool path for manufacturing equipment

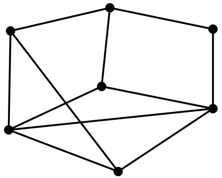




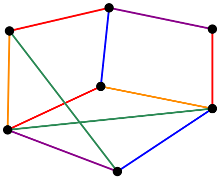
Graph Edge Coloring

An edge coloring of a graph is to assign a color to each edge such that no two "adjacent edges" get the same color

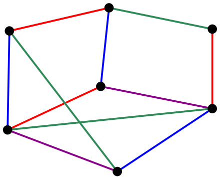
Two edges are adjacent $\longleftrightarrow$ they share an endpoint (incident on same vertex)



Agraph on 7 vertices



An edge coloring with 5 colors



An edge coloring with 4 colors

Want to use the minimum number of colors

$\triangleright$ Minimum colors needed to edge color G is edge-chromatic number

Graphs Applications: Edge Coloring

NFL season scheduling

- n teams in a tournament
- Based on last year's record, each team will play some other teams
- Determine a schedule with as few rounds as possible
- Make a node for each team
- An edge for each game to be played
- Find an edge coloring with minimum number of color

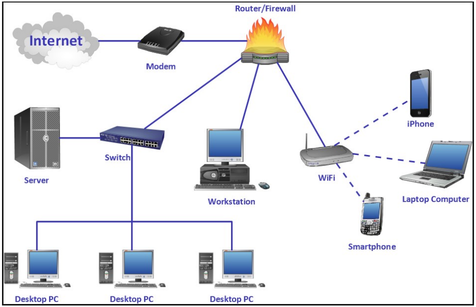
Open Shop Scheduling (time division multi-processing)

- n objects to be manufactured
- Manufacturing object o_i entails performing tasks $t_{i1}, \ldots, t_{ij_i}$ (unordered)
- Each task requires one of non-parallel machines $M_1, \ldots, M_k$
- Make a (multi) bipartite graph [Objects, Machines] edges
- The (o_i, m_j) edge means object i has a task requiring machine m_j
- An edge coloring with minimum number of colors (time slots)

Internet Analysis

The Internet Graph

- Nodes are routers (or sets of routers domains, AS)
- Edges are communication channels



Analyze its structure, find best routes, and many problems

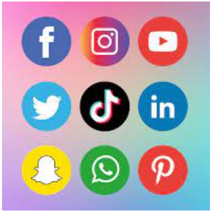
Compute a virtual backbone (a connected dominating set)

Social Network Analysis

Social Network

- Nodes are users
- Edges are social interaction, friendship, acquaintances

Social Network



Online Social Nework

Find important people, for immunization, influencers

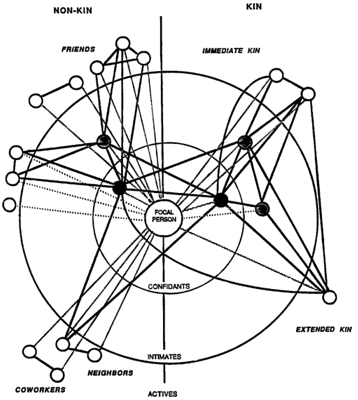
Find communities (for recommendations)

Recommend friends

Extract topics of interests

Social Network Analysis

Network Perspective of Society



An early use of network analysis in sociology. This diagram of the 'ego-network' shows varying tie strengths in concentric circles-Wellman 1998

How the Individuals', communities' and society's behavior is influenced by their social connectivity

Social Influence and Social Selection

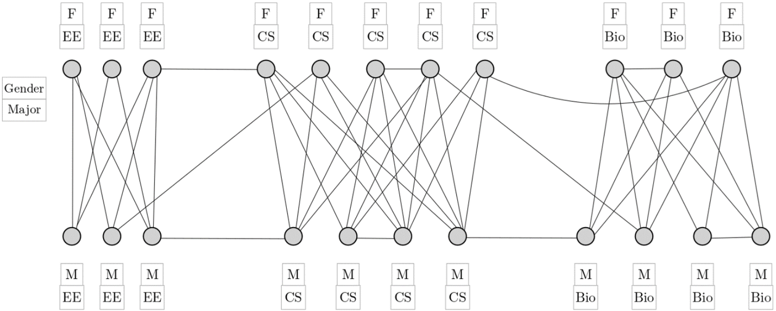
Attribute values and network structure are highly inter-dependent

Two important phenomena in Sociology

- Social Selection: Individual's attributes drive the interaction with others
- Social Influence: Interactions between people shape their attributes

Homophily and Hetrophily

- Homophily: Connections among nodes having same attribute values
  ▷ assortative mixing
- Heterophily: Connections among nodes having different attribute
  ▷ disassortative mixing

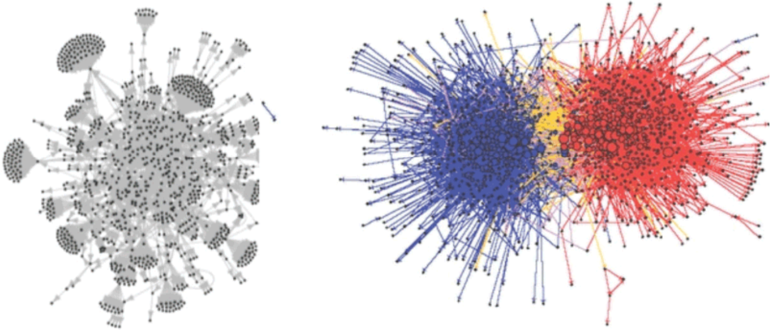


'MAJOR' attribute is homophilic

'GENDER' attribute heterophilic

Network Perspective of Political discourse

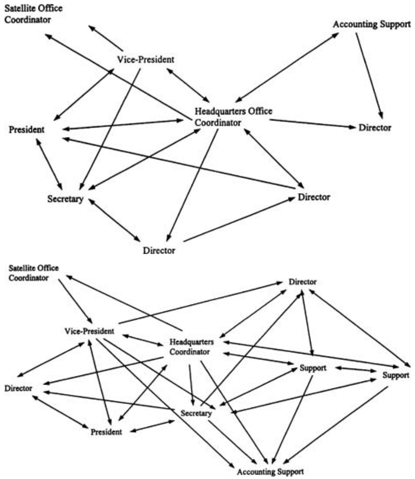
source: Ulicny, Kokar, Matheus, (2010)



(a) Malaysian Sopo blogosphere (b) US political blogosphere (Adamic & Glance, 2005[4])

A visualization of Malysia and US blogosphere (nodes are blogs and edges are links to blogs). Left reveals importance/credibility/popularity of blogs, while the right visual clearly show that bloggers are more likely to link to bloggers with the same party affiliations, forming two dense clusters with little interaction with the other cluster

Communication within an organization

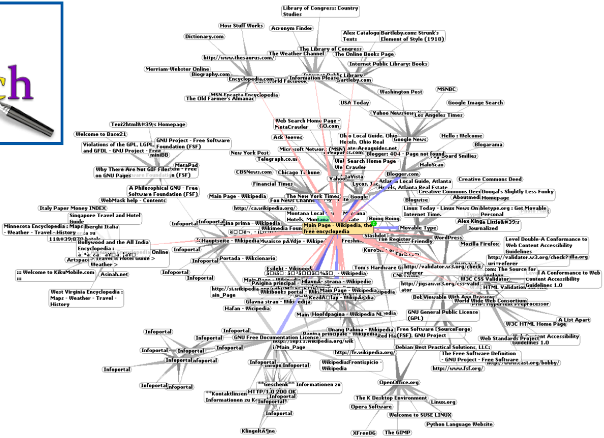


Intra-organization communication before and after implementation of a content management system (Garon et.al. (1997)

Webgraph Analysis

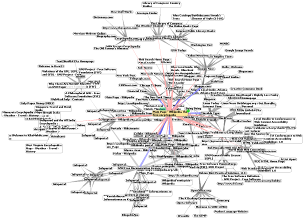
The Web Graph

- Nodes are webpages; edges are hyperlinks
- Analyze its structure
- Search engines ▷ Pagerank and HITS algorithms

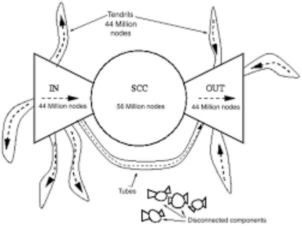


Webgraph Analysis: Web Structure

- Broder et.al. (2000) SCC analysis of webgraph (AltaVista index)
- Study replicated for larger recent webgraphs reveal similar structure



- $\sim 200m$ pages, $\sim 1.5b$ links
- bow-tie structure (macroscopic)
- grouping of SCC's
- CORE: a giant SCC ($\sim 56m$) nodes
- IN: can reach CORE (unidirectional)
- OUT: can be reached from CORE
- TENDRILS:
  - reachable from IN cannot reach CORE
  - can reach OUT not reachable from CORE
- TUBES: both types of tendrils
- DISCONNECTED COMPONENTS



The bow-tie structure of the web (A. Broder et.al (2000))

Graph Analysis: Applications

- Biology (Biological Entities Networks)
  - Discover unknown relationships (disease to disease etc.)
  - Exploratory data analysis and Anomaly detection
  - Protein-Protein Interaction Networks
  - Foodchain in ecosystem
  - Functional network connectivity (FNC) (synchronicity relations between brain parts)
- Geo Information System (Smart Cities)
  - Coverage analysis, traffic flow, congestion estimation, routing
  - Failure impact analysis
- Reasoning (Predictive Maintenance)
  - Predict the next state given the current (and previous states)
  - Compute the probability of sequence of events

Graph Analysis: Applications

- Computer Science: webgraph, Internet, Information dissemination
- Businesses: Analyze and improve communication flow within and between organizations
- Advertisers and Marketers: Figure out the most influential people in a social network and rout message through them
- Security and Law Enforcement: Identify criminal networks from traces (call-logs), find identify key players in such networks
- Banking & Finance: Find unusual patterns in the flow of money across interconnected Banking networks to identify fraudulent transactions, money laundering, terror financing
- Mobile Network Operators: Optimize network structure to enhance QoS, analyze cell towers to ensure maximum coverage