Basic Graph Theory

Mark Ellingham
Department of Mathematics
mark.ellingham@vanderbilt.edu

Notes prepared in collaboration with
Justin Schroeder

****

I. Fundamental graph concepts

II. Random graphs

III. Network motifs and dynamics

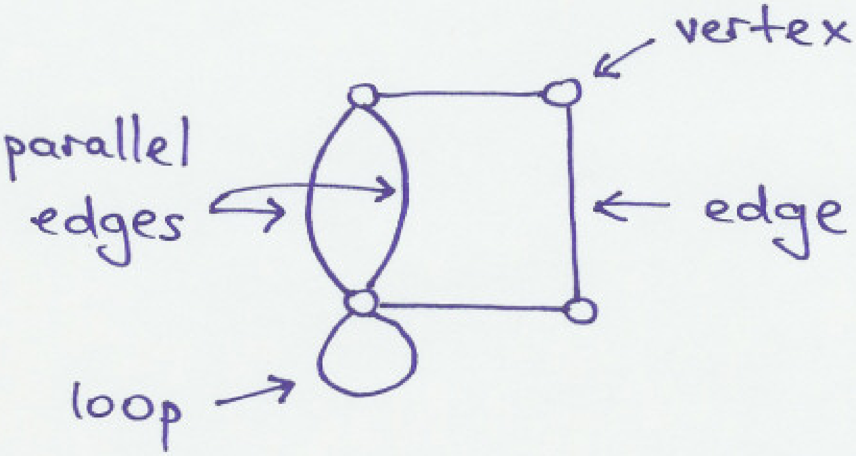
****

I. FUNDAMENTAL GRAPH CONCEPTS

§1 WHAT IS A GRAPH?

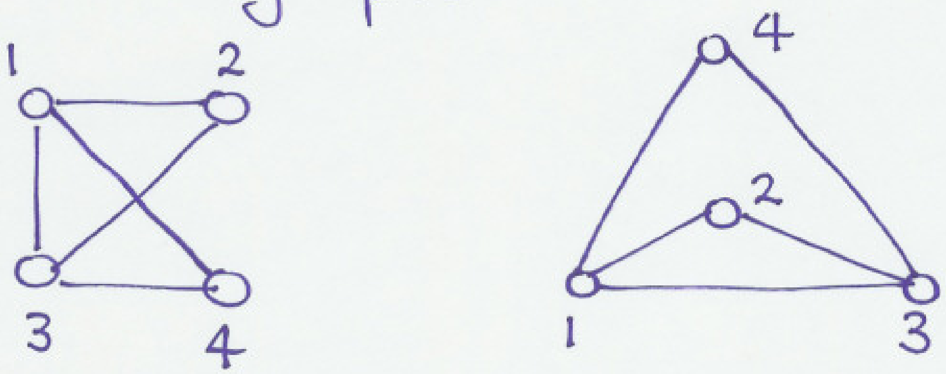
- A **graph** is a (finite) set of vertices together with a (finite) set of edges, where each edge connects two vertices.
- Notation: G = graph, V = vertices, E = edges.
- Example
- Graphs can be **undirected** or **directed**. Additionally, a graph can be weighted.
- Graphs can contain **loops**, which are edges that connect a vertex to itself, and **parallel edges**, which are edges that connect the same two vertices. However, we often consider **simple** graphs that contain no loops or parallel edges.
- Example
- The way a graph is drawn has no particular meaning; only the connections are important.
- Two graphs are **isomorphic** if they are the same after relabeling.
- Example

Graph:

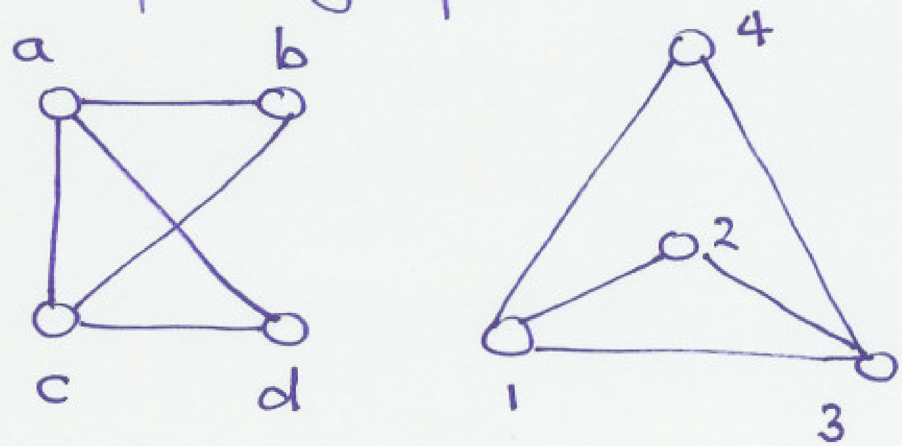


(not simple)

Identical graphs:



Isomorphic graphs:

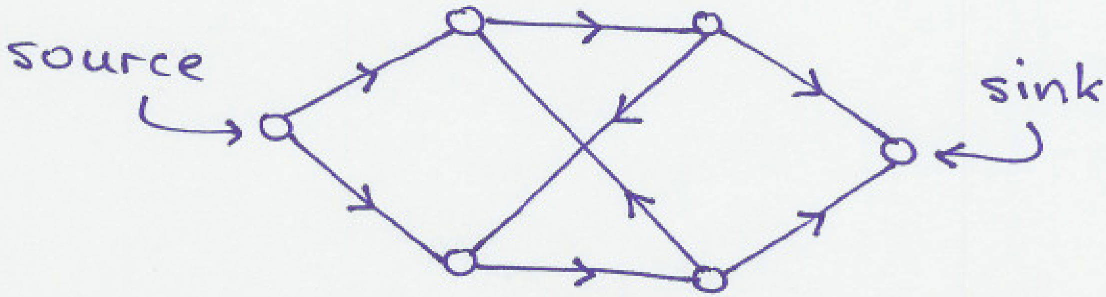


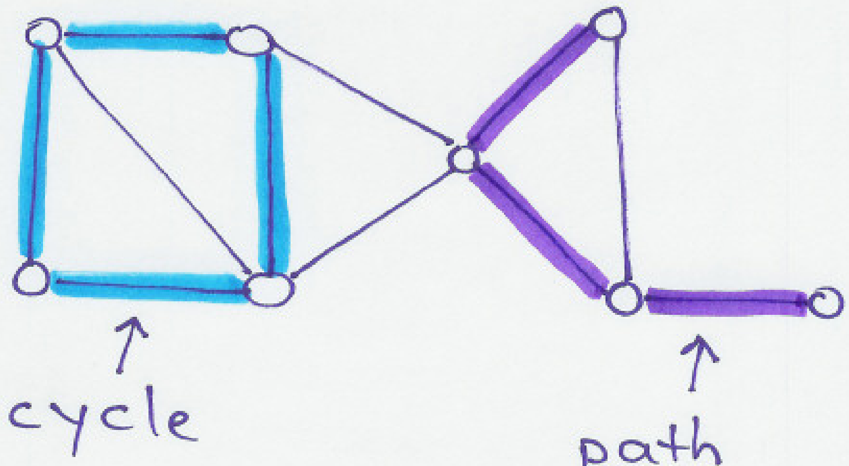
$a \leftrightarrow 1$
$b \leftrightarrow 2$
$c \leftrightarrow 3$
$d \leftrightarrow 4$

§2 CONNECTION CONCEPTS

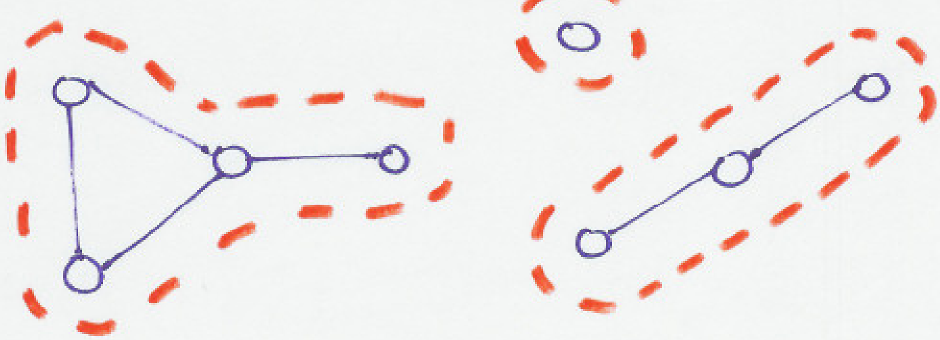
- The **degree** of a vertex is the number of edges incident to it. The **neighbors** of a vertex are the vertices that it shares an edge with.
- In a directed graph, a vertex with all edges directed in is called a **sink**, while a vertex with all edges directed out is called a **source**.
- Example
- A **path** is a sequence of edges that connect two vertices. A **cycle** is a path whose beginning and ending vertex are the same. We often only consider **simple paths** and **simple cycles**.
- Two vertices are **connected** if there is a path between them. A graph is **connected** if every pair of vertices is connected. If a graph is not connected, then we can break it up into its connected **components**.
- Example
- Related ideas for a directed graph: **directed path**, **directed cycle**.
- A graph is **strongly connected** if for every pair of vertices $\{v, w\}$ there is a directed path from v to w and a directed path from w to v.

Directed graph:





Components of disconnected graph:



- Can obtain a new graph by "collapsing" each **strong component** into a single vertex. This graph is necessarily **acyclic**.
- Example

§3 MEASURES OF GRAPH STRUCTURE

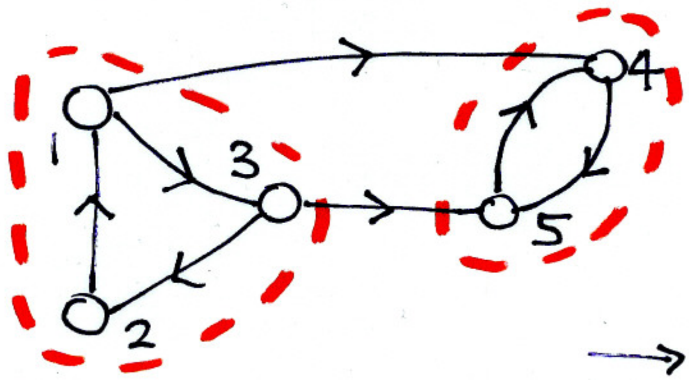
- **Degree distribution**: The number of vertices that have each degree (how many of degree 0, how many of degree 1, etc.).
- Example
- **Average shortest path length**: For each pair of distinct vertices find the length of the shortest path connecting them, then take the mean over all pairs.
- Example
- **Clustering coefficient**: For each vertex v, define its local clustering coefficient C_v as:

$$C_v = \frac{\#\text{ of edges between neighbors of } v}{\#\text{ of possible edges between neighbors of } v}.$$

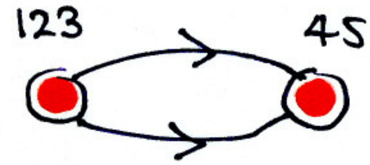
If n neighbors, $\dfrac{n(n-1)}{2}$ possible edges.

The overall clustering coefficient is the mean over all vertices of these coefficients. NOTE: always a number between 0 and 1.

Strong components:



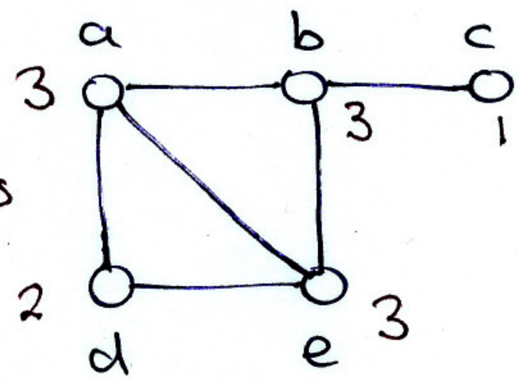
→



acyclic
(no <u>directed</u> cycle)

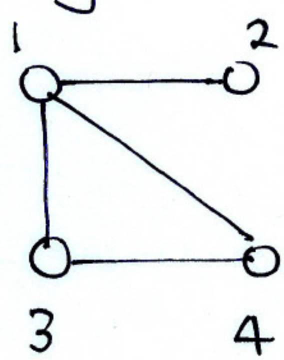
Degree distribution

degrees



1^1	2^1	3^3
c	d	a,b,e

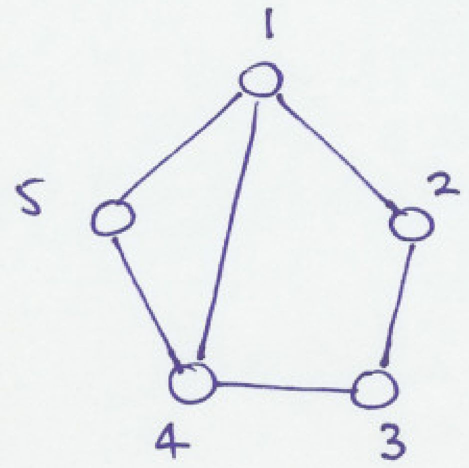
Average shortest path length (ASP)



	1	2	3	4
1		1	1	1
2			2	2
3				1
4				

mean: → $\frac{4}{3}$

Cluster coefficient



$C_1 = 1/3$

$C_2 = 0$

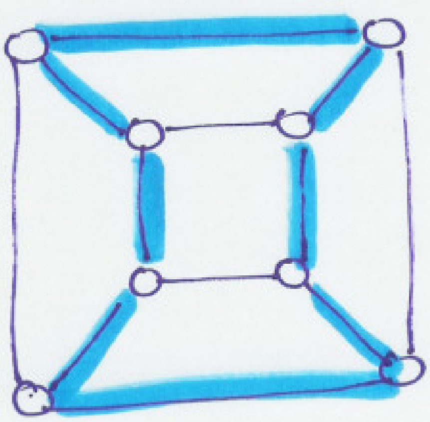
$C_3 = 0$

$C_4 = 1/3$

$C_5 = 1$

Mean: $\frac{1}{3}$

Hamilton cycle



- EXAMPLE
- **Hamilton cycle**: cycle using every vertex exactly once. In weighted graph, may want shortest one.
- EXAMPLE

§4 SPECIAL GRAPHS

- Ring graph
- Lattice (or grid) graph
- Complete graph

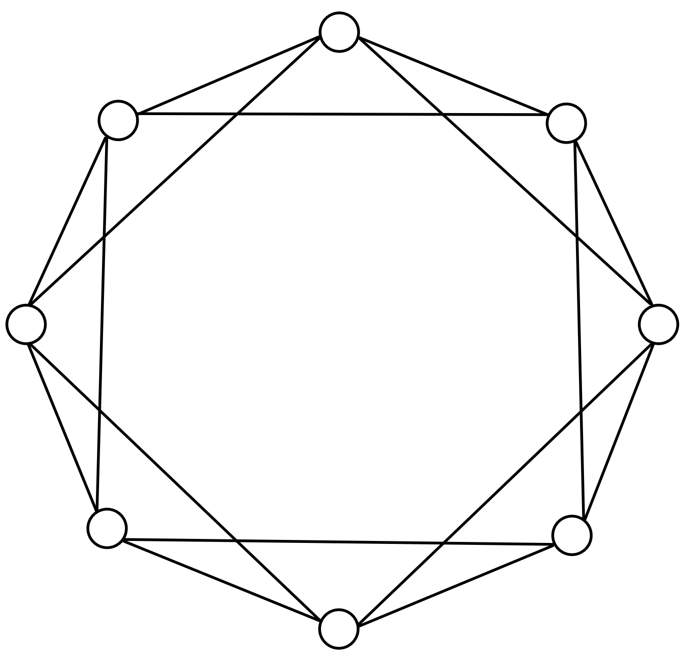


Figure 1: Ring graph.

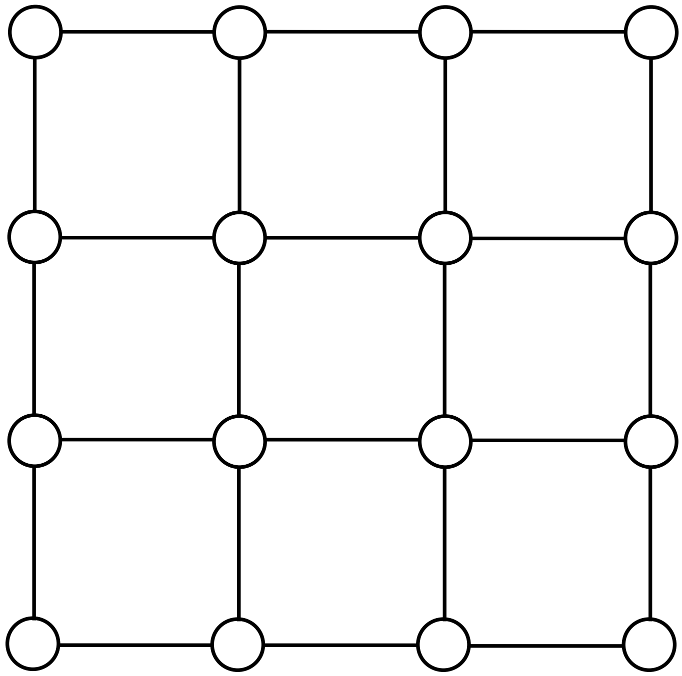


Figure 2: Lattice (or grid) graph.

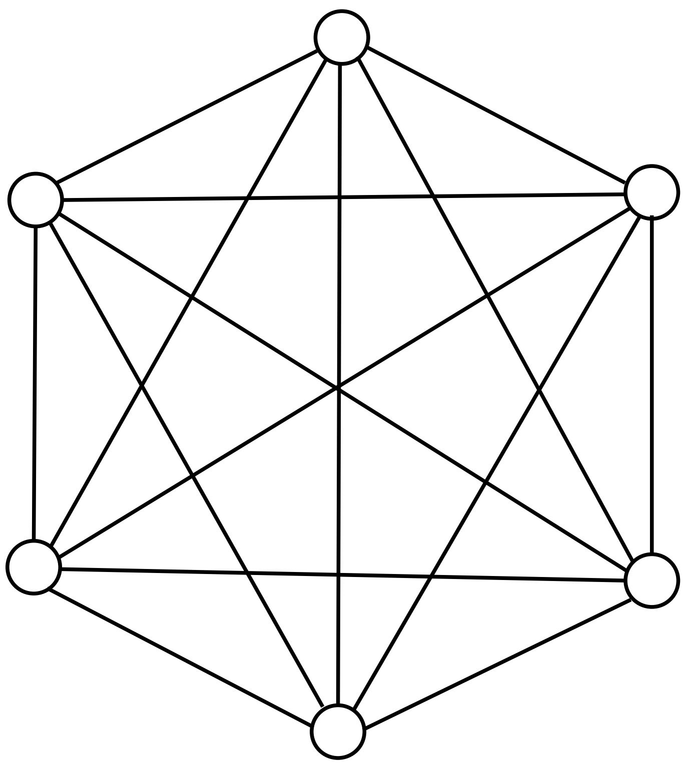


Figure 3: Complete graph.

§5 SOME APPLICATIONS

- Map Color Problem
- Traveling Salesman Problem
- Airline route networks
- Physical computer networks (Internet)
- Virtual computer networks (Web)
- Biological interaction networks

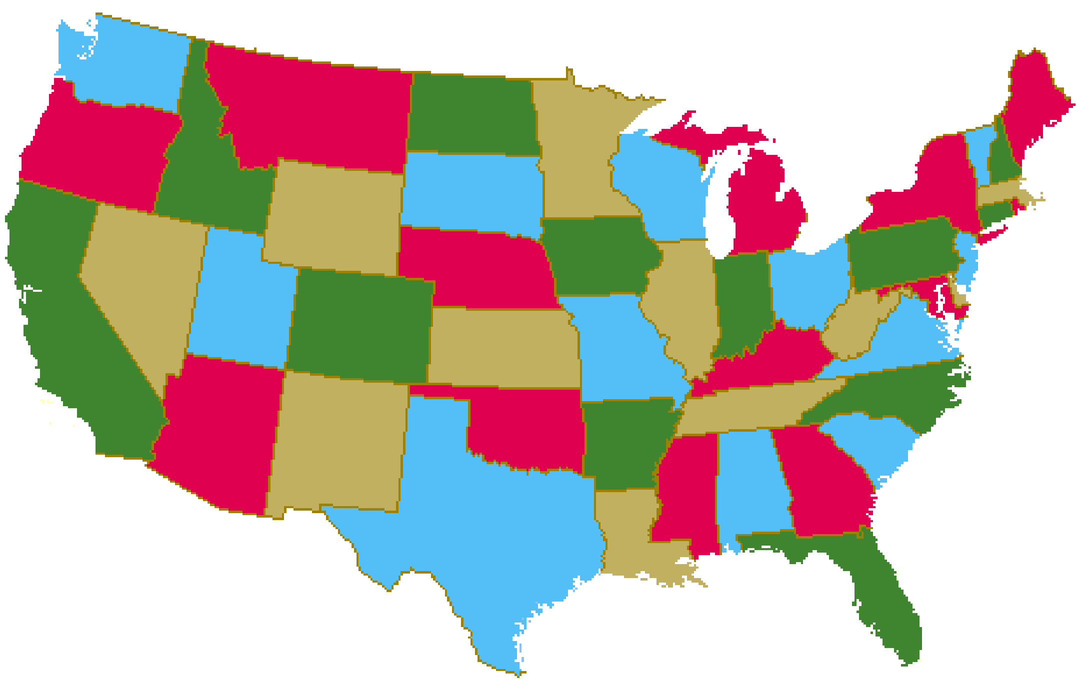


Figure 4: Four-color map of United States.
Robin Thomas, `http://www.math.gatech.edu/~thomas/FC/fourcolor.html`

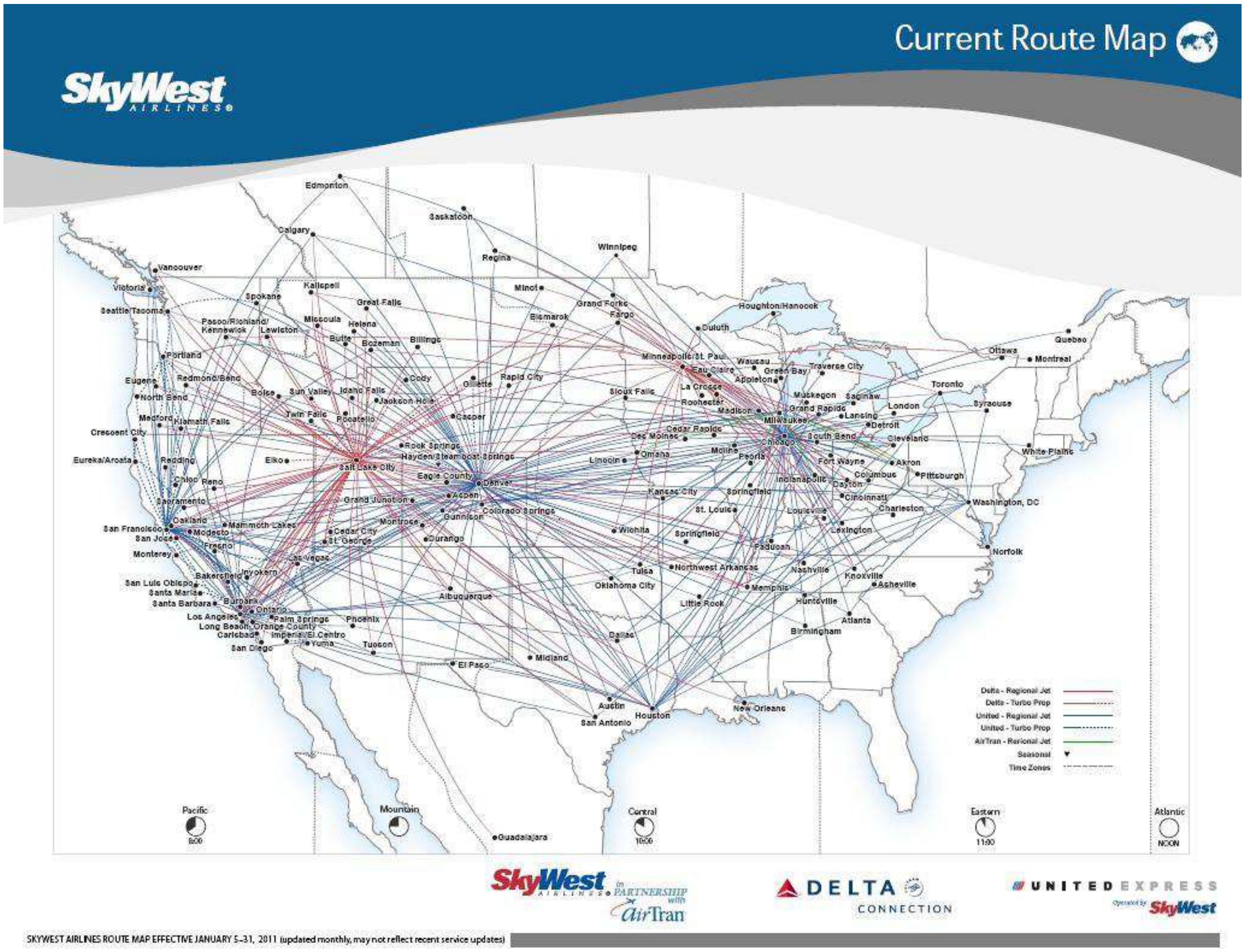


Figure 5: Skywest Airlines routes.

http://www.skywest.com/routemaps/rm_img/SkyWest_RouteSystem%28JAN11%29.pdf

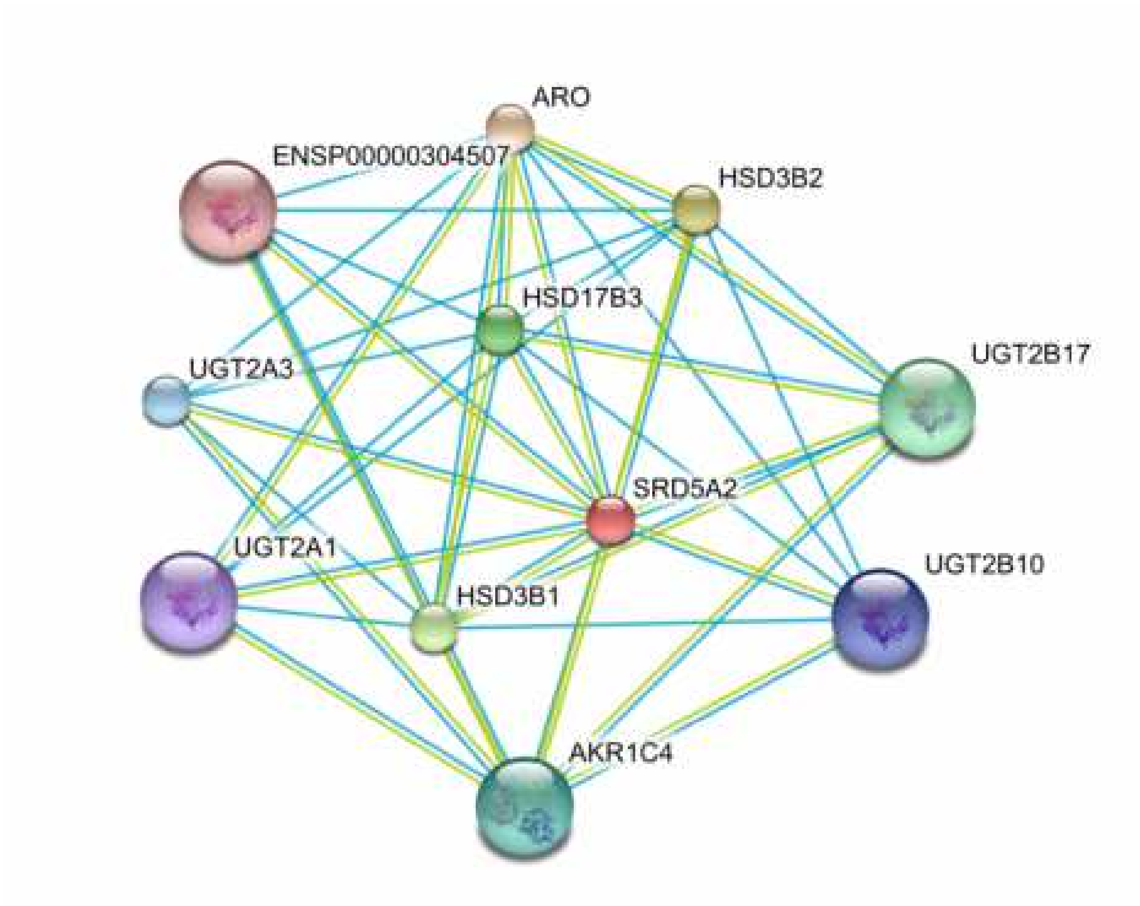


Figure 6: Human protein interactions around SRD5A2.
C. N. Stewart, using STRING 8.0,
`http://stewartgen677s09.weebly.com/protein-network.html`

II. RANDOM GRAPHS

§1 ERDŐS-RÉNYI RANDOM GRAPHS

- A **random graph** is a graph that arises out of some random process.
- Classic construction: **Erdős-Rényi model**
  - Start with a set of n vertices and no edges.
  - For each pair of vertices, add an edge between them with probability p.
  - EXAMPLE
- We can study the expected (average) properties of these graphs (e.g. expected vertex degree is $(n-1)p$).
- We can also study how these properties change as p varies.
- **Major Theorem:**
  - If $p < \frac{1}{n}$, then the graph is almost always disconnected, with no large components.
  - If $\frac{1}{n} < p < \frac{\log n}{n}$, then the graph is almost always disconnected, but will contain a unique giant component.
  - If $p > \frac{\log n}{n}$, then the graph will almost surely be connected.

Erdős-Rényi model

$p = \frac{1}{2}$:

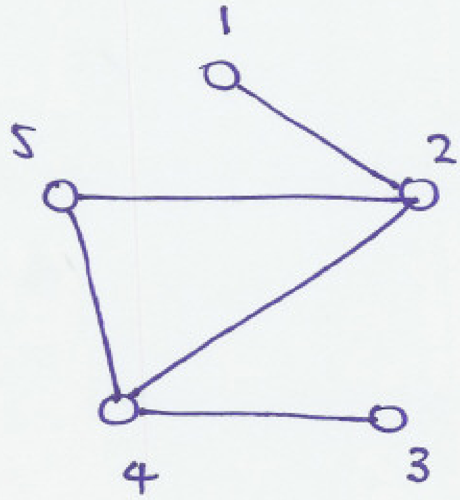
12: T 13: H 14: H

15: H 23: H 24: T

25: T 34: T 35: H

45: T

T = yes, H = no



§2 SMALL-WORLD NETWORKS

- Recall the definitions of average shortest path length (ASP) and clustering coefficient (CC).
- A graph with short ASP and high CC is called a **small-world network**.
- Real world examples
  - Social networks
  - Gene networks

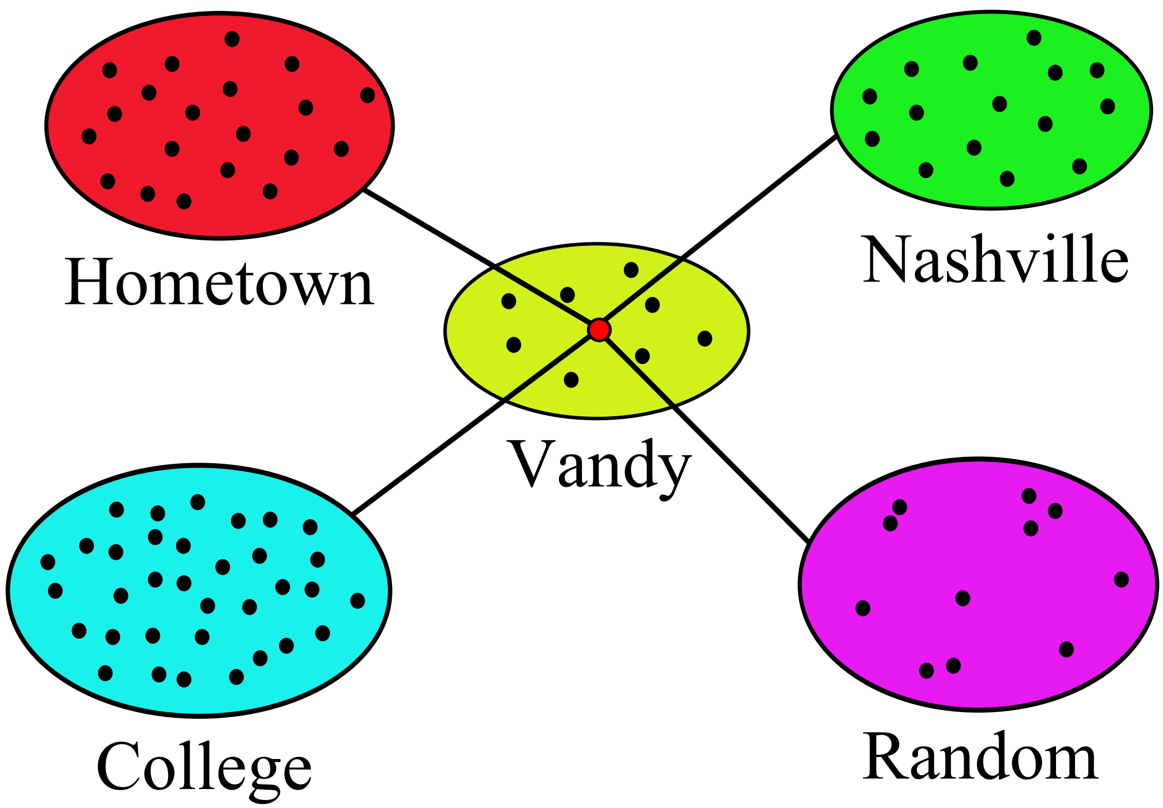


Figure 7: Justin's social network.

- Method to construct such graphs: **Watts-Strogatz model**
  - Start with a ring graph on n vertices with even degree k.
  - For each edge, rewire its clockwise end with probability p (don't change other end).

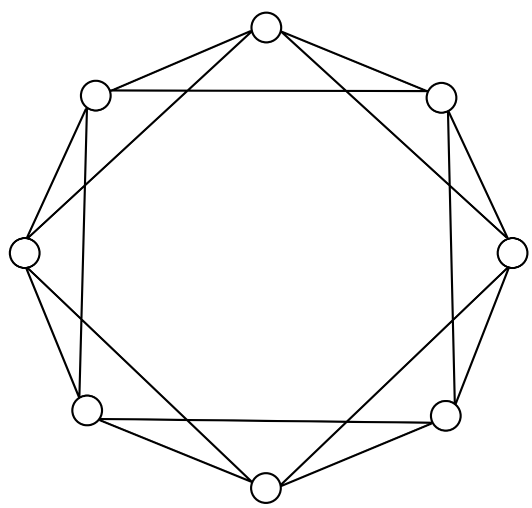


Figure 8: Ring graph (again).

- Behavior
  - If p is close to 0, then this graph is close to being regular.
  - If p is close to 1, then this graph is close to the ER model.
  - If p is in the middle, then this graph has small-world properties.

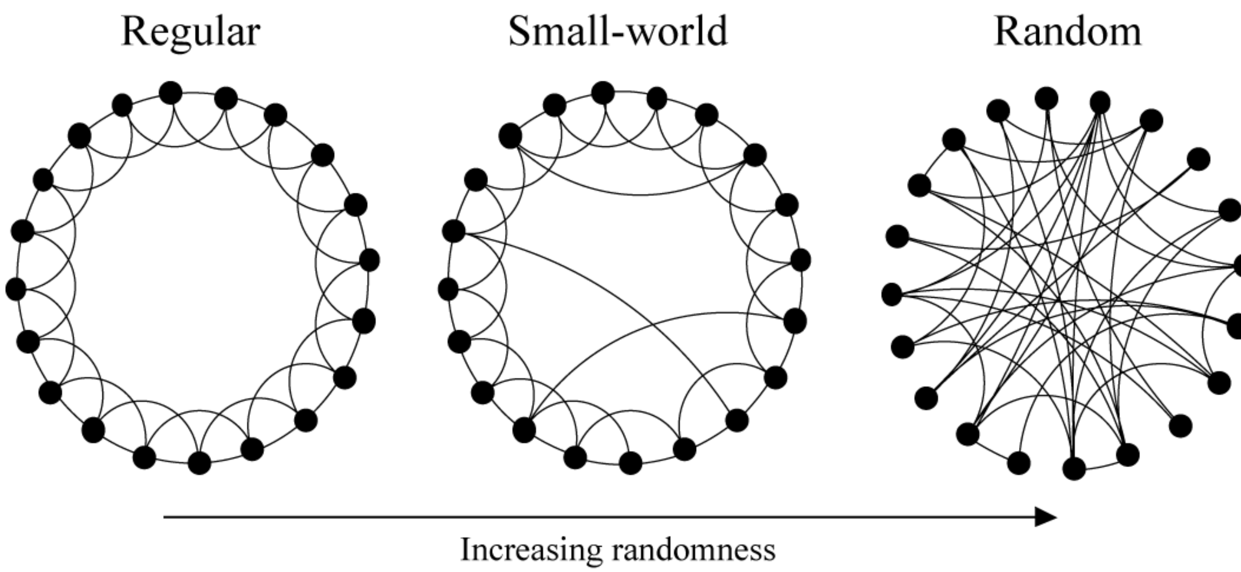


Figure 9: Watts-Strogatz model.
D. J. Watts and S. H. Strogatz, Nature 393 (1998) p. 441.

§3 SCALE-FREE NETWORKS

- Recall definition of degree distribution.
- If the degree distribution satisfies {# of vertices of degree d} $\sim d^{-k}$ for some positive k, then it is a **power-law distribution**.
- A graph with a power-law distribution is called a **scale-free network**.
- Real world examples
  - Internet
  - Protein networks
- Two methods to construct such graphs

* Preferential attachment: **Barabási-Albert model**

  – Start with a graph on n vertices such that the degree of each vertex is at least one.

  – Add new vertices one by one, connecting them to each existing vertex with probability based on the degree of that vertex. A vertex with higher degree has a higher probability of having an edge with the new vertex.

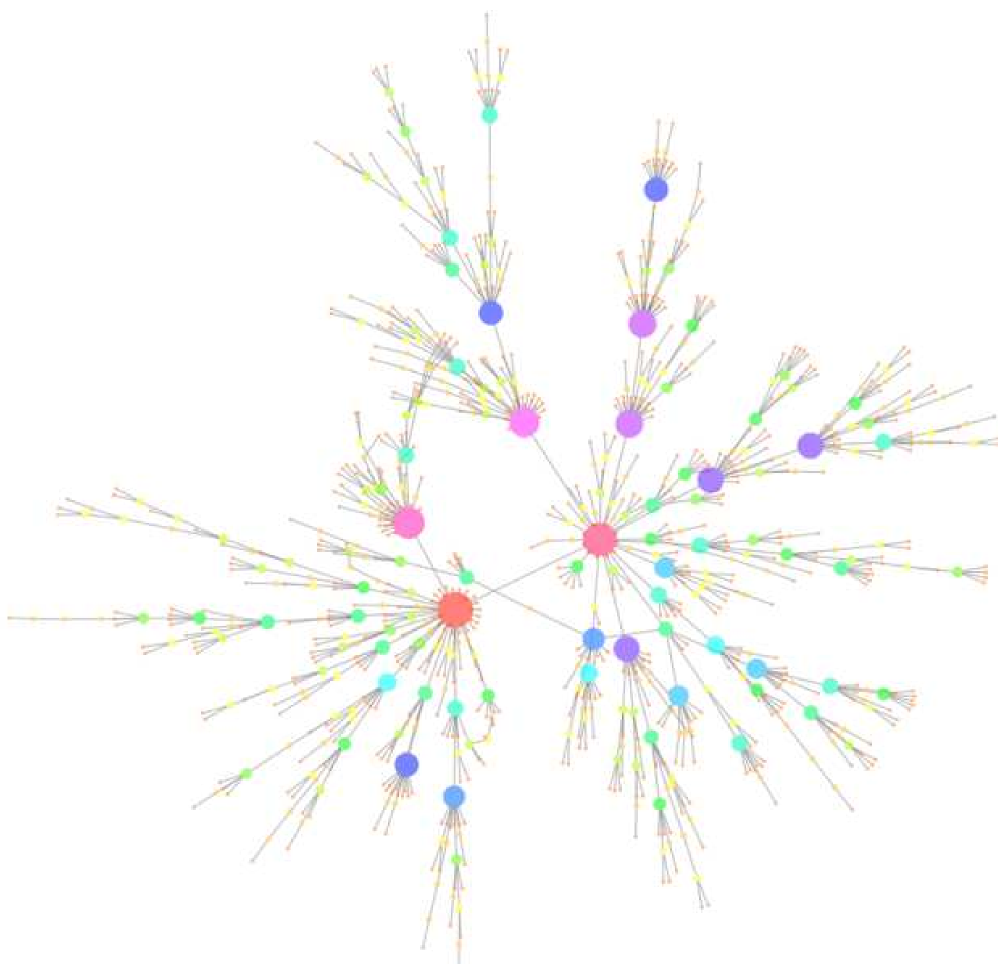


Figure 10: Barabási-Albert model
Keiichiro Ono, `http://commons.wikimedia.org`

* Vertex duplication (Chung et. al, 2003)
  – Start with a graph on n vertices with any number of edges.
  – Randomly select a vertex v and add a new vertex v'. For each neighbor w of v, add an edge between v' and w with probability p.

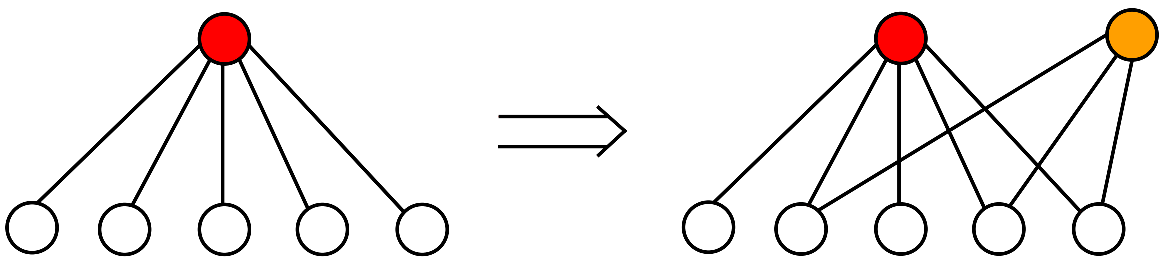


Figure 11: Vertex duplication model.

§4 COMPARISON OF MODELS

- In the ER model, we first had the method, then derived the properties of the resulting graphs. For the other models, we first had some desired properties, then devised methods for constructing graphs with these properties.
- Average shortest path length
  – Most random graphs have short ASP. In fact, the ER and WS models have similar ASP's.
  – Most scale-free networks, in particular the BA model, have shorter ASP's than the ER and WS models.

- Cluster coefficient
  - The ER model has a low CC.
  - The WS model has a high CC, close to $\frac{3}{4}$.
  - CC behavior varies in scale-free networks. Therefore some scale-free networks are additionally small-world networks, but many are not.
- Degree distribution
  - The ER model has a binomial distribution. In other words, most vertices have degree close to the average, with very few vertices of high or low degree.
  - The WS model has a degree distribution heavily concentrated around starting parameter k.
  - The BA and duplication models both have power-law distributions in the form of d^{-k}. For the BA model k is around 3, while for the duplication model $k < 2$.
  - The ER and WS models can be disconnected by random removals of vertices ("attacks"). The scale-free networks are less susceptible to random attacks, but targeted attacks on vertices of large degree (**hubs**) can destroy the network.

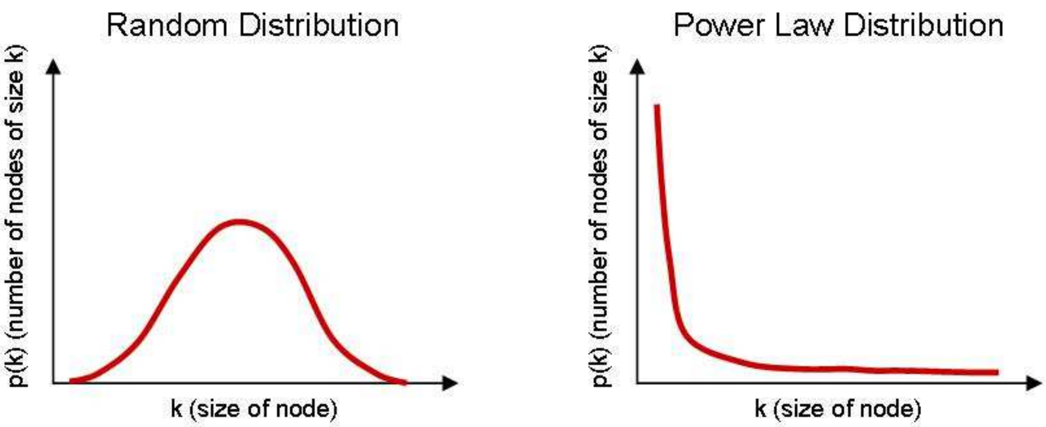


Figure 12: Comparison of degree distributions.
Sebastian Pritchard, `http://bastiatblogger.blogspot.com/2010/06/value-investing-levy-flights-and-tight.html`

III. NETWORK MOTIFS AND DYNAMICS

§1 NETWORK MOTIFS

- We have previously studied the global properties of graphs (ASP, CC, etc.). What can we say about local properties?
- A **motif** is a small substructure that appears frequently throughout the network.
- Four commonly found motifs
  - **Feed-forward loop** - behavior predetermined for a given input
  - **Feedback loop** - behavior can be affected by output
  - **Single-input module** - one input controls many outputs
  - **Dense overlapping regulon** - many inputs overlap to control many outputs
- The SIM and DOR motifs can be found with standard algorithms. Detecting loops is considerably more difficult.

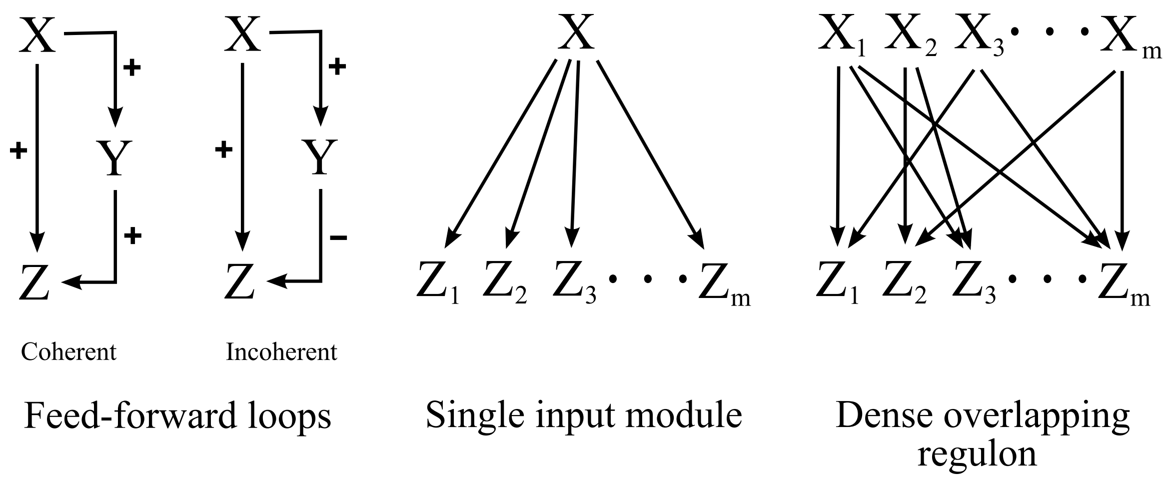


Figure 13: Some common motifs.

§2 FEEDBACK STRUCTURES

- A feedback loop consists of a directed cycle
- Switching behavior can be provided by a positive feedback loop

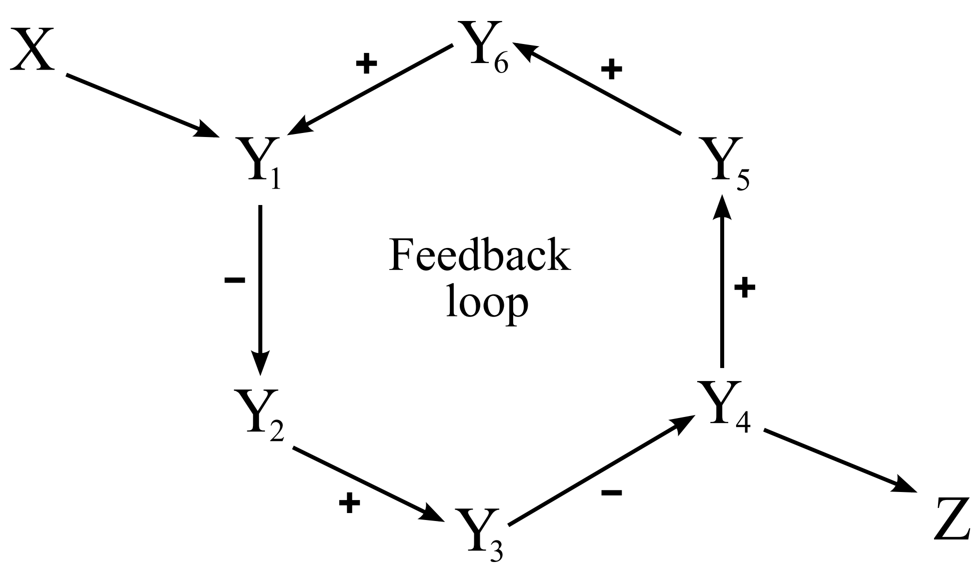


Figure 14: Example of a feedback loop.

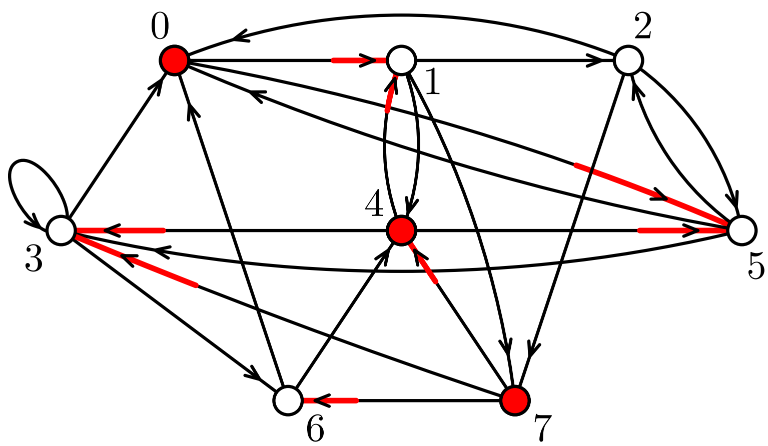
- Finding such a loop is often challenging, but this problem can be simplified in a couple of ways.
  - Eliminate sinks and sources (never in directed cycle). If result is relatively small, check directly for positive directed cycles.
  - Subdivide each positive edge into two edges. Positive cycles become even cycles. There is a polynomial time algorithm to see if an even cycle exists.

§3 NETWORK DYNAMICS

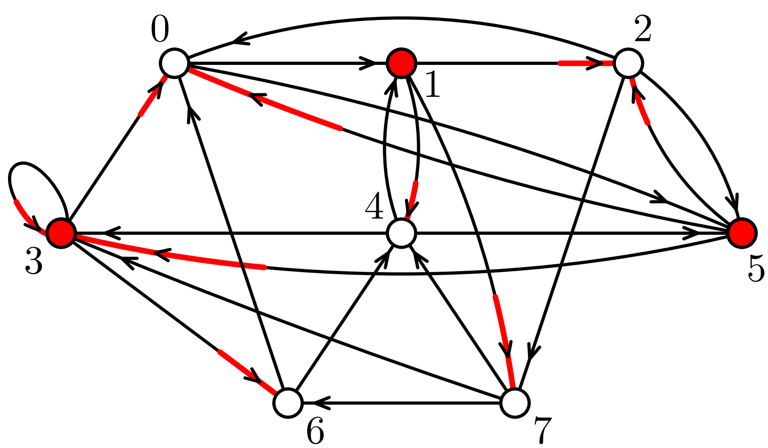
- In addition to the properties we have studied thus far, we are often interested in how a network changes over time. These **network dynamics** help us understand the behavior of the network.
- One model: **Boolean network** (Kauffman)
  - Each vertex is "off" (0) or "on" (1). This determines the **state** of the network, given by a binary number.
  - For each time step, the **evolution** of the network is determined by a set of rules. These rules depend on the current state, and are often simply a threshold.

– In the example below, we are using a threshold of 2. So for a vertex to be "on" (red) at time t, it must have at least 2 "on" vertices pointing to it at time $t-1$.

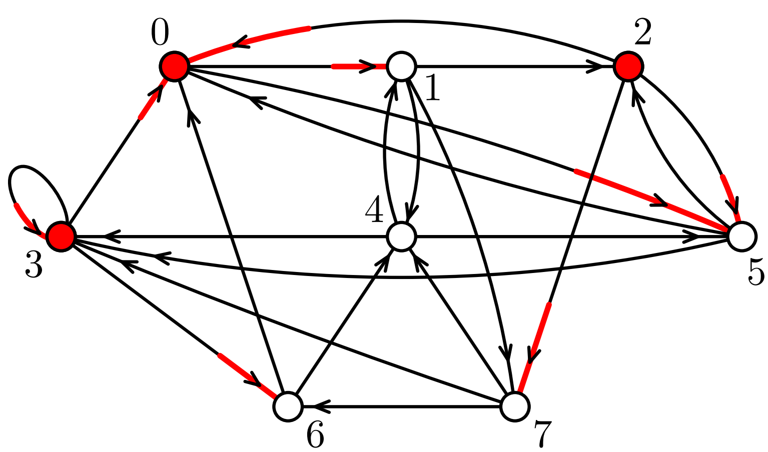
Step 0



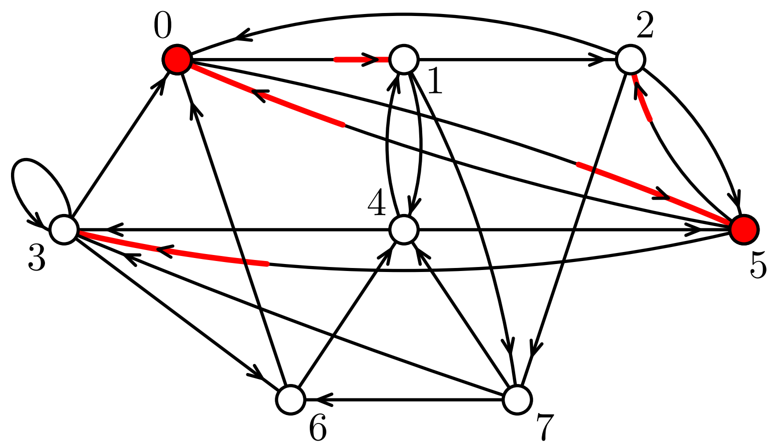
Step 1



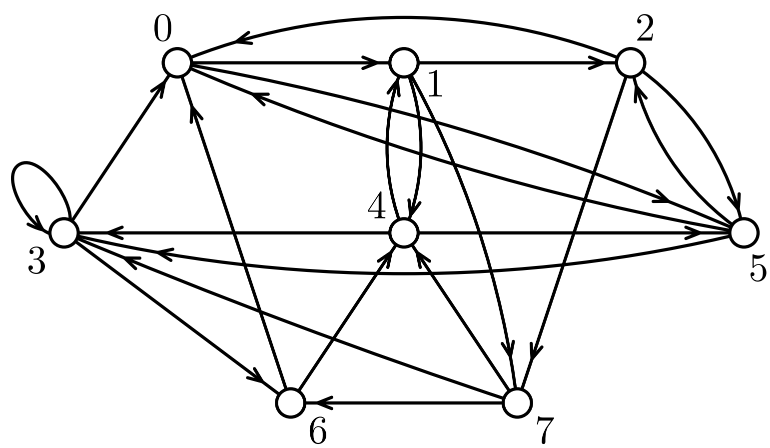
Step 2



Step 3



Step 4

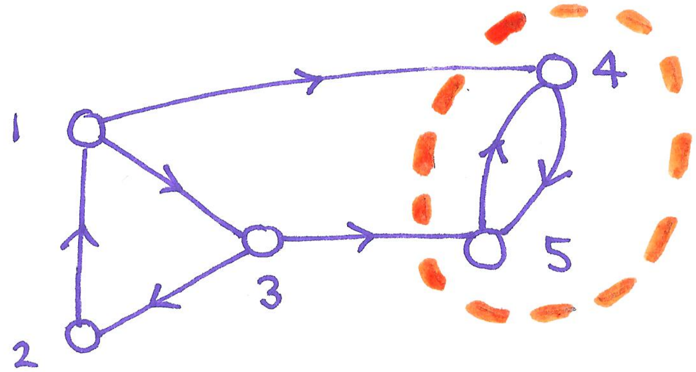


- State graphs
  - There are many possible states for a given network. In a boolean network on n vertices, there are 2^n states.
  - We can form a graph with these states as vertices, and an edge from state S to T if applying the rules to S yields state T.
  - An **attractor** is a state, or cycle of states, that you never leave once you reach it. (E.g., Step 4 in example.) The states that lead to an attractor are the **basin** of that state.

§4. MEASURING CONNECTIVITY IN GRAPHS

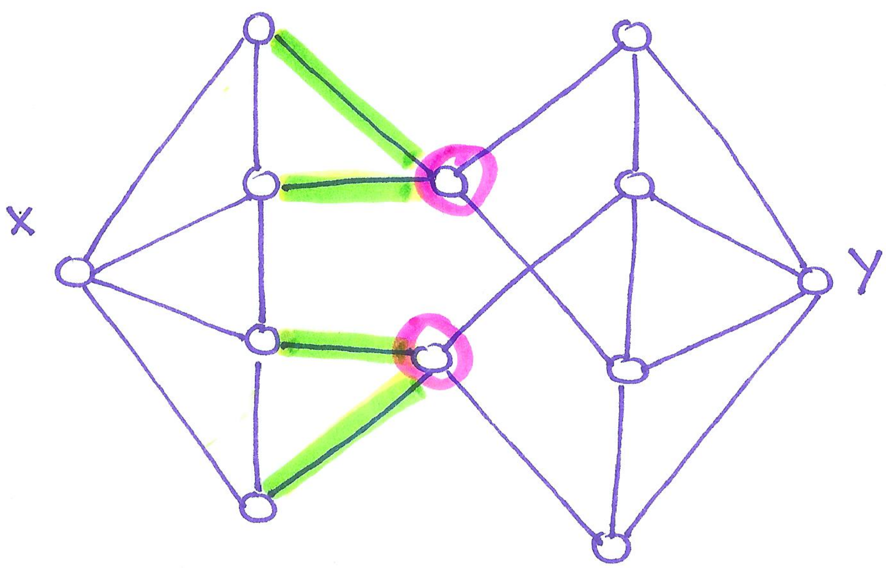
- A **cut** is a collection of edges that, if removed, will disconnect the graph.
- Let X and Y be a subset of vertices of a directed graph. The set of edges from X to the other vertices, which we denote $\delta^+(X)$, is called the **outcut** of X. Similarly, all the edges from the other vertices to X, $\delta^-(X)$, is the **incut** of X.
- A directed graph is **strongly connected** if $\delta^+(X) \neq \emptyset$ for every proper nonempty subset X of V.
- EXAMPLE
- We can quantify **connectivity** in graphs (directed or undirected), identify **bottlenecks**, in various ways:
  - **edge connectivity** from x to y: minimum number of edges whose removal destroys all xy-paths;
  - **overall edge connectivity**: minimum over all pairs of vertices;
  - **(vertex) connectivity** concepts: remove vertices instead of edges.
- EXAMPLE

Not strongly connected:



$X = \{4,5\}$

$\delta^+(X) = \phi$



xy - edge connectivity = 4

xy - (vertex) connectivity = 2

- Connectivity k from x to y is equivalent to the existence of k 'disjoint' xy-paths:
  - edge connectivity $\leftrightarrow$ paths that share no edges;
  - vertex connectivity $\leftrightarrow$ paths that share no vertices except their ends.
- **Practicality:** We can compute connectivity efficiently in many situations by using **network flow algorithms**:
  - vertex or edge cuts,
  - weighted or unweighted,
  - directed or undirected.