Graph Thoery and Social Networks

Nathan Feldman

DRP Summer 2014
University of Maryland

September 2014

Outline

Outline

Definition of a Graph

- A graph consists of a triple of information (V(G), E(G), ϕ_G)
- V(G) is a non-empty set called the vertices
- E(G) is a set called the edges
- ϕ_G is called the incidence function which associates to each edge an unordered pair of vertices
- A graph is often represented by a picture with points as vertices and line segments as edges with endpoints of each edge attached to the vertices that are assigned to that edge.

Example Graph

- Here is an example with V(G)={1, 2, 3, 4},
  E(G)={$e1, e2, e3, e4$}
- In this example $\phi_G(e1) = (1, 2), \phi_G(e2) = (1, 3), \phi_G(e3) = (1, 4), \phi_G(e4) = (3, 4)$
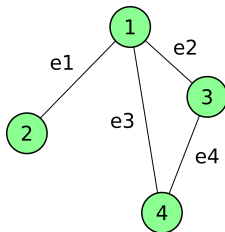
Outline

Weighted Graph

- A weighted graph is a graph with a number, called a weight attached to each edge

- This weight in applications can represent distance, the strength of a connection between nodes, or the cost to transfer something from place to place

- We can now introduce the notion of distance between two nodes

- The distance of a particular path from one node to another is just the sum of the weights

- The distance between two nodes is then defined to be the minimum distance over all paths between the nodes

Dijkstra's Algorithm

- One interesting problem in the theory of weighted graphs is the shortest distance problem
- Dijkstra's Algorithm generates the shortest path to any other node from a fixed node
- For those who know about computation time, this particular algorithm runs in time $O(|V(G)|^2)$

Outline of Dijkstra's Algorithm

- You start off by including only the vertex you wish to start at and label each adjacent vertex with its distance from the vertex in question

- At each step connect the vertex with the smallest label along the path with the labeled distance and call that vertex A

- Then, to each vertex adjacent to A, you compare the current label on that vertex with the sum (the distance of that vertex to A) + (the label on A)

- you replace the label with the smaller of the two values and repeat the process by then connecting the vertex with the smallest label

Outline

Basics of Social Networks

- Social networks can be described as a graph of individuals who all come in contact with each other.
- This can take a form similar to Facebook where individuals "friend" each other in a process known to all of us on social media.
- Networks can also form in real life at a college where individuals form a large-scale network based on club involvement and students belonging to multiple clubs

Social Networks as Graphs

- There are five main characteristics that describe a graph that will be used to represent a social network:

- Firstly, all edges are undirected, without one vertex pointing to another.

- Secondly, edges are unweighted. All "bonds" within a network are assumed to be equal.

- Thirdly, the graph must be simple with no vertices connecting to themselves at all or other vertices multiple times.

- Fourthly, the maximum size of the set of edges must be $\binom{n}{2}$, where n is the number of vertices, corresponding to the number of edges present in a complete graph

- Finally, the graph of a single network must be connected. There must be a path between any two vertices of the graph.

Outline

1. Background Information
   - Graph Theory Basics
   - Weighted Graphs

2. Social Networks and Topic Modeling
   - Social Networks
   - Graph Metrics
   - Topic Modeling

Graph Metrics - characteristic path length

- The characteristic path length of a graph is the median of the means of the shortest path lengths connecting each vertex of a graph to all other vertices.

- This means to calculate $d(v,j)$ for all vertices j in a graph given a vertex v, and for all vertices v, find the mean of $d(v,j)$. Then take the median of all the means.

- In social networks, networks with low characteristic path lengths can have information spread much more rapidly to the entire population within the network. On the other hand, high characteristic path lengths can make information take longer to spread.

Graph Metrics - clustering coefficient

- The clustering coefficient of a vertex is defined as the number of edges connected to that vertex over the number of possible edges.
- For a graph, the clustering coefficient is defined as the mean of the clustering coefficients for the individual vertices.
- In social networks, it can be useful to target individuals with high clustering coefficients to spread information throughout the network.

Outline

1. Background Information
   - Graph Theory Basics
   - Weighted Graphs

2. Social Networks and Topic Modeling
   - Social Networks
   - Graph Metrics
   - Topic Modeling

Topic Modeling
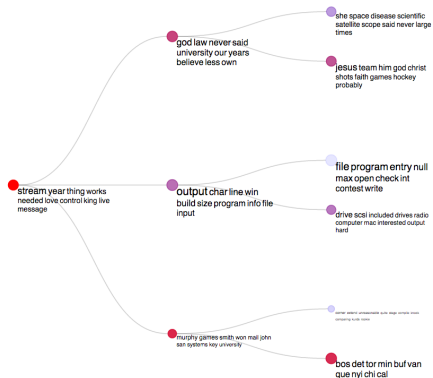
A Brief Discussion of Topic Modeling

- These topic models relate individual users or user-generated content to particular topics with various probabilities.

- The models were constructed using machine learning algorithms to slowly create more accurate models.

- These topic models are meant to help an outsider gain more information about a particular user.

For Further Reading I

📕 J.A. Bondy, U.S.R. Murty
Graph Theory with Applications.
Elsevier, 1976.

📄 P. Jahnichen.
Finding and Analyzing Social Networks in unstructured web log
data using probabilistic topic modeling.