**Project eCSite Unit:** Social Networks and the Tipping Point

**Content Areas:** computer science, data analysis & probability, measurement, science & technology

**Grade Level:** Secondary (9 - 12)

**Computational Thinking Connection:** This unit uses data collected from social networking websites (e.g., Facebook) to allow the students to prove or disprove claims made about social network structure in the book “The Tipping Point” by Malcom Gladwell. In order to do this, the students must: (1) form testable hypotheses (2) grok basic graph theoretic algorithms (i.e., mean average shortest path/degree of separation) (3) understand how to analyze and present data to answer hypotheses, including summative statistics, and fundamentals of statistical distribution shape and structure. Overall this tends to reinforce the lesson that computer science and statistics are tools which allow you to do things like data analysis to answer (personally and scientifically) interesting questions about topics as diverse as social structure and word of mouth epidemiology.

**Prerequisite Knowledge:** The unit presumes that the students are secondary level and are in the process of discussing social epidemics, probably in the course of a sociology or marketing course. The students should read at least Chapter 2 (“The Law of the Few”) of “The Tipping Point” by Malcom Gladwell.

**Time Required:** 2-3 classes

**Related Lessons / Activities:** The unit works about like this:

1) Students are assigned to read chapter 2 of “The Tipping Point”  
   (An abridged reading can also be done in class)
2) An in-class or take-home activity gets students thinking/discussing social networks  
   (assignment provided)
3) Students are assigned to use the “Facebook Data-Miner” tool to upload their data  
   (assignment provided)
4) The resulting data is extracted from the Data-Miner by the teacher and prepared for use in the next two activities  
   (instructions provided)
5) An in-class activity in a computer lab walks the students through using Fusion Tables to analyze distributions and hypotheses in from the data.  
   (activity provided)
6) An in-class activity in a computer lab walks the students through analyzing the structure of the data using the Gephi graph visualization tool or This can be done on a projector, and discussed in class.
7) Optionally: privacy implications w/r/t corporate use/misuse of data should be discussed.
Contents:

1) Data investigation activity (4 parts)
2) Instructions for using the data-miner software
3) Instructions for doing basic statistical analysis and hypothesis testing

Links:
Gephi, an open-source graph visualization application:
http://gephi.org/

Google Fusion Tables:
http://www.google.com/fusiontables/Home/

Software to go with this unit:
http://systems.cs.colorado.edu/~caleb/gk12_tipping_point_social_nets.tar.bz2

Other Notes: N/A

Contributors: Caleb Phillips and John Williams

Supporting Programs: NSF and GK-12 eCSite program
Social Network Data Investigation Activity

In this activity, we're going to analyze our own class' Facebook social network to try to observe some aspects of social networks described by Malcom Gladwell in the *Tipping Point*. In particular, we're going to try to prove the existence of connectors, mavens and salespeople within our own social group.

PART 1: Thinking About Sociology and Social Networking

Instructions

This activity contains a series of discussion questions to get you thinking about sociology and online social networking tools. You will get into groups to discuss and answer the questions. For each of your responses, record each group member’s contribution. Afterward, we will discuss as a group.

Background

In case you haven’t yet read chapters 1 and 2 of Malcom Gladwell’s, *The Tipping Point*, here are a couple of definitions:

*Connectors* are the people who "link us up with the world ... people with a special gift for bringing the world together."[^5] They are "a handful of people with a truly extraordinary knack [...] for making friends and acquaintances".[^6] He characterizes these individuals as having social networks of over one hundred people.

*Mavens* are "information specialists", or "people we rely upon to connect us with new information."[^4] They accumulate knowledge, especially about the marketplace, and know how to share it with others. According to Gladwell, Mavens start "word-of-mouth epidemics"[^9] due to their knowledge, social skills, and ability to communicate. As Gladwell states, "Mavens are really information brokers, sharing and trading what they know".

*Salesmen* are "persuaders", charismatic people with powerful negotiation skills. They tend to have an indefinable trait that goes beyond what they say, which makes others want to agree with them.

Questions

1. Do you use social networking websites, if so, which? If not, why?

2. How often do you use these websites?

3. Do you think your online social network is similar or different from your “real world” social network? Explain.
4. Do you think the “law of the few” applies to online social networks?

5. How might these “special few” people use, e.g., Facebook, differently than the rest of us?
   a. Connectors:
   
   b. Mavens:
   
   c. Salespeople:

6. What are positive uses of these websites that you’ve seen?

7. What are negative uses of these websites that you’ve seen?

8. What can these social websites tell us about contemporary society? How are they changing the way social relationships work? Is this good or bad? Why?

9. What sort of information do people post on social networking websites?

10. Realizing that these networks are run for profit by companies, do you trust the companies to use your information wisely? Why or why not?
11. What are some concrete ways do you think this information could be abused?

12. Can you think of ways this information can be used for good?
PART 2: Using the Facebook Data-Miner

Instructions
Follow these instructions to extract some data from your profile and put it in our database. It will only take a minute or two. This will not violate your privacy or store data that could get you in trouble in any way. All this application will store is:

- The number of interests you have (i.e., movies, books, etc.)
- The number of groups you belong to
- The number of times someone has “liked” or “commented” on a post that you've made in the last 30 days
- The number of posts that you've made in the last 30 days.
- Your friend list, and the same numbers listed above for each friend (if their privacy settings allow it)

Upload Your Data

First off, if you don't have a Facebook account, you'll need to make one. It isn't hard, you can do it here: http://facebook.com.

Next:
1. Go to: http://smallwhitecube.com/fb/
2. Click on the “Click Here to Login” link. A Facebook login window will pop up (make sure you aren't blocking the popup)
3. Login to facebook and give the application permission to access your account. The popup window will go away and you'll be logged in.
4. Click “Update Your Data”. This will take a minute and will count to nine.
5. If you get an error, then click “Update Your Data” again. If there’s still an error on the second try, you can ignore it. You're in the database now!
6. You can click logout now. You're all done.

You will be graded based on whether there is an entry in the database with your name.

Recruit Your Friends

The more data we can get, the more cool stuff we can do. So, ask your friends (at BHS or anywhere) to follow the instructions above and upload their data too. The student who gets the most other people to upload their data will win a gift card to Illegal Pete's.
PART 3: The Law of the Few and Power Law Distributions

Instructions

This activity will walk you through loading and analyzing data from our social network (extracted from Facebook). While you work through the activity, there are some questions you need to answer. It should be obvious what you need to answer, but we've made them bold to help you find them.

You need to answer questions in *full sentences* and with some thought or you will only get partial credit. One word answers and/or answers copied and pasted from online sources will get zero credit.

When you are done, email the completed document to: caleb.phillips@colorado.edu with the subject: “BHS Activity 1”. If your email address doesn’t make your name obvious, then put it somewhere in the document or email. If you don’t have an email account, you can print it out and hand it in.

Background

A histogram is a plot that shows the number of times something occurs (the frequency). For instance, here’s a plot that shows the number of Black Cherry trees of a certain height.

![Histogram of Black Cherry Tree Heights](image)

It shows that the most popular height for Black Cherry trees is approximately 78 feet. And that most cherry trees are between 70 and 80 feet tall. We might call this the “distribution” of Black Cherry tree heights. This is a “normal” distribution because it is shaped like a bell, with a clear peak at the center and sides that slope down.

Another type of distribution, more interesting for us, is a “power law”, which looks like so.
Let’s say that this plots the frequency of the population of cities. In this distribution, all of the weight is to one side, which means that most cities are small, but since there is a long tail off to the right, it also means that a few cities are very large. This is also known as the 80-20 rule. For instance, 20 percent of the people, have 80 percent of the wealth. If we looked at this plot as a plot of the number of people having some amount of wealth, you can see that most people have little wealth (they are on the left), but a few people have lots of wealth (they are on the right). As we discuss the “law of the few” below and how it relates to social networks, this concept will be important...

**The Law of the Few in Our Own Social Network**

Today we’re going to see if we can find evidence of the law of the few in our own social network...

1. Go to this link to see the data we collected from Facebook last week:  

2. This is a big “table”, where each line is a person in the extended social network of this class. How many people are in the network? Is that more or less than you would expect?

3. Now, let’s see if we can find evidence of the law of the few. We don’t have good data showing how many friends each person has since Facebook privacy settings block this, but we do know how many people in *this class* that each person in the list is friends with. Let’s look at that data.

4. Click View -> Aggregate. Then select “nconn” next to “Aggregated by:” and finally click Apply.

5. This shows the number of the people in the network having each number of friends within the class. Let’s plot this data.

6. Click Visualize -> Line. Then choose “Count” for the Y-axis.

7. Does this look like a power law? Why or why not?
8. How many people in our network are friends with only one person in the class? How many people are friends with 15 people in the class?

9. Can you explain how (or why) someone might be “friends” with more than half the people in our class?

10. Does this graph support the law of the few? Why or why not?

11. Now, let’s try to find Mavens in the data. Since Mavens collect information, it seems likely they would have many interests relative to most people. Let’s look at that. Click Aggregate, then uncheck “nconn” and check “ninterests”. Then, click “Apply”. Select “Count” for the y-axis as before.

12. Does this look like a powerlaw? Why or why not?

13. How many people have only 1 interest? How many people have more than 400 interests?

14. Due to changes in facebook privacy settings, it’s difficult to get counts of people’s comments or “likes”. However, if we could get this data, how would you expect the distribution to be similar or different from that of “interests” given above?

15. Do you think the people that appear to be Mavens, Connectors, and Salespeople in facebook, would also be Mavens, Connectors, and Salespeople in “real life” (i.e., as Gladwell describes these roles)? Why or why not?

16. With the rest of the time, search around on Google for other examples of Power Law distributions. List three things here that also have been shown to follow a Power law:
PART 4: Finding Patterns and Structure in Social Networks

Instructions

This activity will walk you through loading and analyzing data from our social network (extracted from Facebook). While you work through the activity, there are some questions you need to answer. It should be obvious what you need to answer, but we've also made them bold to help you find them.

You need to answer questions in full sentences and with some thought or you will only get partial credit. One word answers and/or answers copied and pasted from online sources will get no credit.

When you are done, email the completed document to: caleb.phillips@colorado.edu with the subject: “BHS Activity 2”. If your email address doesn’t make your name obvious, then put it somewhere in the document or email. If you don’t have an email account, you can print it out and hand it in.

Background

A network (also called a “Graph”) is a way of visualizing how things are connected. Here’s a picture of a simple network of 6 “nodes”. For our purposes, we’ll imagine the nodes are people and that the lines between them indicate that those people are friends (on Facebook). Hence, Person 6 is friends with person 4, but not with person 2. And, person 1 is friends with both person 5 and person 2.
Check out this, more complicated, network:

Betweenness describes how “in between everyone” a given person is. This is another way of saying that they are a “person who ties many other people together”. In the figure above, the people with the most highest “betweenness centrality” are blue and the people with the lowest “betweenness centrality” are red.

Closeness centrality is similar, it simply says how “far” someone is from everyone else. A person with a small closeness centrality is basically friends with everyone, and a person with a high closeness centrality may be several “friends-of-friends-of-friends” away from everyone else.

Download and Install The “Gephi” Program

1. Go to http://gephi.org/users/download/
2. Click Mac OS X
3. Click Download Gephi Alpha for Mac OSX
4. In the bottom right, where downloads go, click on gephi-08-alpha.dmg, an icon “gephi” will appear on the desktop. Double click this icon.
5. Then click “gephi”, and click “open” to open the gephi program
6. Download the all.dot and common.dot files from:
   http://systems.cs.colorado.edu/~caleb/all.dot
   http://systems.cs.colorado.edu/~caleb/common.dot

Loading in the Facebook Network Data

1. Open the common.dot file
2. Leave the import settings as they are and click “ok”
3. A bunch of black dots and lines in a square should appear. This is all the people in the shared social network of the class that know at least 2 people (i.e., all the “common friends”). But, this is difficult to draw anything meaningful from.
4. Click the “partition” tab on the top left and then on “Nodes”

5. Click the little green arrows to refresh the drop down list. Then click “Label” in the drop down list and then click “apply” to color the graph.

6. The nodes will all get random colors. The folks in the class will each have separate colors and the folks not in the class will all get the same color. Stil it’s a bit hard to make sense of this plot.

7. Click the “T” in the bottom-center of the screen to turn on node labels. This will put a number over each person in the class showing how many friends there have. You can probably find yourself...

8. Find where it says “Layout” the center left and choose “Force Atlas” and click “Run”

9. When it seems to have stabilized and stopped moving, you can click “Stop”

A First Look at Our Common Social Network

The file common.dot has only people that at least 2 people in the network know (folks that only one person know have been filtered out). Let's try to understand this, (our) common social network...

1. First, in the top right of the screen find the number of nodes and edges. The number of nodes is the number of unique people in the network. The number of edges is the number of unique friendships. Write down those numbers here:

2. Is there anyone in the class you can see that appears to be a “connector”? If so, how many friends do they have? Where are they positioned relative to everyone else?

3. Are there any people not in the class that seem central to the graph and seem to know many people within the class? Who do you think these people are?

4. Can you identify people that may be friends (outside of class)? How?

5. Looking at this network, can you think of how you might use the available information to “suggest friends” the way Facebook does?
Digging deeper into the Network Structure

1. Now, find where it says “Average Degree on the right. Click Run.

2. The average degree is the average number of (common) friends anyone in this network has. What is that number? Did you expect it to be higher or lower?

3. After you click run, there will also be a report that shows the degree distribution. Does this seem to follow the 80-20 (“power law”) rule? In other words, do a small number of people have the most connections, while most people have a few connections?

4. Now click Run next to “Network Diameter”, this will calculate the network diameter and the average path length. It will also show plots of the “betweenness centrality distribution” and “closeness centrality distribution”. Let’s look at these one at a time.

5. The average path length is the average “Milgrim” (or Kevin Bacon) distance from any person in the network to any other person. Write down that number here. Is it higher or lower than you expected? Why or why not?

6. The network diameter is the maximum “Milgrim” distance from any person in the network to any other person. Write down that number here. Is it higher or lower than you expected? Why or why not?

7. Now look at the report and the distribution of closeness centrality and betweenness centrality. These are known as “centrality metrics” and provide a way to understand how important a given person is to a given (social) network. Betweenness centrality is how “in between” a person is between all other people. Closeness centrality is how “close” a person is to all other people.

8. Describe the shape of these distributions. Do they seem to obey an 80-20 (power law) rule?
9. Would a connector have a high or low betweenness centrality? What about closeness centrality? Using these “metrics” does this network seem to have some connectors?

**Wrapping Up**

1. Given all of the above, do you believe that our own Facebook social network data supports the “law of the few” described by Malcom Gladwell in the tipping point?

2. How do you do you think a social networking tool like facebook facilitates/enables “social epidemics” of the sort Malcom Gladwell describes?

3. Optional: If you have time, load all.dot and play around with the complete, unfiltered social network data. What sort of interesting things can you pull out of this data?

Want to learn more? *This Wikipedia page gives a good overview of social networking research:*  
ALTERNATIVE PART 4: Discussion Questions

If a computer lab is not available to allow the students to run part 4, it can be completed as a class. In this scenario, the teachers loads up gephi and the questions below are used for the basis of class discussion.

1. First, in the top right of the screen find the number of nodes and edges. The number of nodes is the number of unique people in the network. The number of edges is the number of unique friendships. Write down those numbers here:

2. Is there anyone in the class you can see that appears to be a “connector”? If so, how many friends do they have? Where are they positioned relative to everyone else?

3. Are there any people not in the class that seem central to the graph and seem to know many people within the class? Who do you think these people are?

4. Can you identify people that may be friends (outside of class)? How?

5. Can you identify any other interesting patterns or outliers in the data?

6. Looking at this network, can you think of how you might use the available information to “suggest friends” the way Facebook does?

7. The average degree is the average number of (common) friends anyone in this network has. What is that number? Did you expect it to be higher or lower?

8. After you click run, there will also be a report that shows the degree distribution. Does this seem to follow the 80-20 (“power law”) rule? In other words, do a small number of people have the most connections, while most people have a few connections?

9. The average path length is the average “Milgrim” (or Kevin Bacon) distance from any person in the network to any other person. Write down that number here. Is it higher or lower than you expected? Why or why not?

10. The network diameter is the maximum “Milgrim” distance from any person in the network to any other person. Write down that number here. Is it higher or lower than you expected? Why or why not?

11. Describe the shape of the closeness centrality and betweenness centrality distributions. Do they seem to obey an 80-20 (power law) rule?
12. Would a connector have a high or low betweenness centrality? What about closeness centrality? Using these “metrics” does this network seem to have some connectors?

13. Given all of the above, do you believe that our own Facebook social network data supports the “law of the few” described by Malcom Gladwell in the tipping point?

14. How do you think a social networking tool like Facebook facilitates/enables “social epidemics” of the sort Malcom Gladwell describes?
How to Use the Application

This unit involves the use of a data-miner application that mines Facebook for data which is then analyzed in the class. This software runs on a server. When working, after a user has authenticated (with Facebook) and logged into the application, they see something like this:

The admin view includes a link for “Generate Report” on the right but students only see “Update Your Data”. When students login and click “Update your Data”, they see:
When logged in as the admin (the admin account is configured as part of the software installation) you can download the stored data in a couple different formats:

- Download Common-Friends Graphviz Graph
- Download Entire Graphviz Graph
- Download User Information as CSV
- Download Friend Data as CSV

Hello, Caleb Phillips!

This is a little Facebook application which extracts some data from your Facebook profile and stores it in a database. It stores: (1) your friends list, (2) counts of your posts, comments, and likes, (3) counts of your friends' posts, comments, and likes. It doesn't store any personal data except names, genders, and these aggregate counts. The application was developed to be used in a classroom setting, and specifically for a sociology class at Boulder High School in Boulder, Colorado.
By downloading as a Graphviz graph, the data can be analyzed with the Gephi software which enables discovery of graph structure, visualization, and some basic statistics. For instance:

By downloading the data as a CSV, the instructor can do more data analysis (ideally with the help of the students). The script analysis.R provides some hypothesis testing where appropriate. The included C program allows you to calculate average shortest path between all pairs of nodes (using the Floyd Warshall Algorithm). Here is an example of the output of these analysis routines. Each figure speaks to one of the numbered hypotheses from the lecture nodes above. The ones that are immediately testable from the data we can get out of facebook are:

H1: Number of friends is a powerlaw (finding connectors)
H2: Average path length to all other nodes is a powerlaw (finding connectors)
H4: Amount of content generated is a powerlaw (finding mavens)
H6: Amount of positive feedback is a powerlaw (finding salespeople)

The instructor should talk about why these are testable in a straight-forward way and why the others are not. Some hypotheses aren't easy to test because of what Facebook chooses to make accessible via the API. This can start a discussion on data privacy.
The concept of powerlaw distributions and heavy tails should be tied intimately to Gladwell's law of the few. Students must really understand what a distribution means and how to interpret the shape.

Starting with H1: Number of friends is a powerlaw.

This figure is simply a histogram and it isn't very conclusive.

A smoothed probability density plot of the same data shows that number of friends appears to be roughly normal with a mean around 500. Again, not helpful for testing the hypothesis.
However, if we plot number of incoming edges (i.e., not how many friends people have total but how many they have in our data-set)...

And this, looks like a powerlaw!

Or, viewed another way:
Other hypotheses are less easy to prove, but still can offer an opportunity for some discussion...

<table>
<thead>
<tr>
<th>Number of People</th>
<th>Mean</th>
<th>Standard Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1420</td>
<td>2.83</td>
<td>0.2</td>
</tr>
<tr>
<td>5065</td>
<td>3.33</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Not so good...

**H4|Mavens: Amount of Generated Content is a Powerlaw?**

**H2|Connectors: Average Shortest Path Length is a Powerlaw?**
Nor is this, but is still worth discussing. Amount of data available for these last two is inherently limited by Facebook's privacy settings.