Questions for text Chapter 8

1. a) Who are important individuals within the network?

This is perhaps the most fundamental question one can ask of a social network. It is useful to know who are the most active participants, which may be a good indication of importance (though not necessarily in the case of email, due to spam; hopefully this will be filtered out in the analysis). I'd like to know which people I spend most of my time emailing; actual results may not match my expectations.

b) What collaborative activities are individuals engaged in?

It could be useful to understand the group activities in which the user participates, such as mailing lists for classes or interest groups; perhaps the user would like to contribute more to certain discussions and less to others. It also may be the case that the user initiates email conversations with groups of recipients, such as sending out mass emails of dubious value, and would benefit from a visual reminder of activity that may not be a good use of time. Again, I'd like to know if my expectations about this match reality.

c) What are the relationships between subgroups?

Characterizing the communications between subgroups can identify opportunities for new collaborations and synergies. Groups that have little or no communication might benefit from bridge-building, perhaps facilitated by the person being studied (me, in this case). It is also possible for the person being studied to gain greater understanding of their current role in fostering communication between groups; perhaps they have a high betweenness centrality, or perhaps communications are mediated more by another person.

2. a) Aside from me, the most important person in the network judged by edge weight (i.e., number of email messages exchanged) is my coworker for my Graduate Research Fellowship, Jason, from whom I've received 43 messages and to whom I've sent 27 messages in the past month. This makes sense, as we collaborate closely on our work projects and we were preparing for a presentation last week so traffic was even higher than usual. Also involved in this work effort are Allen Malony, my sponsor, and Gwen Frishkoff, the co-PI on the project, and they have very high edge weights as well. Examining the edge weights reveals that I receive a large amount of messages from mailing lists, particularly grads-mail and dept, and some from my class mailing lists (cis677 and cis607).

Sorting by betweenness centrality reveals that Allen Malony is a very important link between all the people in my CIS 630 class and me. There's no mailing list for this class, and he emails the class members by including us all in the "To:" field of his messages. I have only emailed a few of my classmates directly, so most of the connections I have to the class members are only through him, giving him a high betweenness centrality score (1870). Also important in this area are others who routinely send messages to mailing lists, including Star, Cheri, Jim Allen, and Jan, and the mailing lists themselves.

b) There are clear subgroups visible from the network visualization, including those for individual classes I am taking and for my work. Of course, they are not independent (unconnected) groups, as many people overlap between the groups. This is particularly the case for work and CIS 630, as Allen Malony is a common element in both and there are quite a few edges involved in those.
c) Relationships between subgroups seem to be mediated by certain individuals with a high level of betweenness centrality, as mentioned in section (a). And as mentioned in section (b), there is quite a bit of overlap between groups, which is not surprising considering that I am taking courses with many of the same students. There is no overlap, however between my CIS 677 contacts and any of my other groups. Though I believe at least one of my CIS 677 classmates is also in another class of mine, he has not sent messages to me or to the cis677 mailing list so there is no link visible on the network map.

3. As the section about analyzing Enron emails makes clear, SNA can play a useful role in investigating fraud, but by itself it does not prove fraud or criminal activity. It can help show who was in communication with whom, and when the dataset being analyzed is filtered appropriately this can give an indication of corporate activity in response to certain situations or influences (when filters are applied to keywords, or time periods, for instance). SNA can provide insight into the credibility of individuals: if someone says that they were not kept in the loop about a certain situation but the email record shows they were emailed many times about that situation, then their credibility naturally is called into question.

In order to actually find someone responsible for criminal activity, it would be much more useful to see the contents of email messages to understand exactly what was communicated. So, SNA can point out patterns of communication and give investigators a place to focus their efforts, but the actual contents of email messages will likely be more incriminating than communication patterns in general, as the book points out.

There is no such thing as private email, unless it is encrypted on the client computer before sending and decrypted on the recipient's computer only when reading and never stored unencrypted on either end. This is possible to implement, but very uncommon in civilian settings. In normal email systems, messages reside in plain text and are stored on or pass through multiple systems during composition, transmission, and reading. There has been a trend to move email to cloud systems such as gmail, and even users who don't use such systems are typically moving to IMAP, in which messages are stored on the server rather than downloaded from server to client and deleted from the server, as is usually the case with POP systems. In any case, messages can be intercepted or accessed at multiple places, whether illegally or through court orders. Corporate governance rules often dictate that corporate communications must be archived and preserved for quite a long time, so any corporate email users should assume that their messages will be available to authorized (and potentially unauthorized) parties for many years.