I. Setup and Dataset

I ran NodeXL from a virtual machine running Windows 8.1, using Parallels [1]. I used NodeXL to import 15 months worth of email from Microsoft Outlook, corresponding to the amount of time I’ve been at UMD. Both the Cc line and Bcc line were used in calculating edge weights, and the maximum email size was restricted to 2Kb so that the data would load. This yielded a dataset with 310 nodes representing email addresses and 567 edges representing emails sent between addresses. Nodes will remain unlabeled for privacy’s sake.

II. Network structure

I found that using the Harel-Koren Fast Multiscale algorithm created the most understandable initial display of the network, represented here as a digraph with edges shaded and sized by edge weight:

![Email network displayed with Harel-Koren](image)

**Fig. 1:** Email network displayed with Harel-Koren

Fig. 1 centers on my email address, the node at the center. The cluster of nodes immediately around my address with high edge weights represents the people I am in most frequent contact with (the big blue arrow is my mom). Going outward from the center, there are two interesting features: Clusters of highly connected nodes (cliques),
representing academic and social communities, and “fans,” single nodes connected to many nodes. Fans correspond to email lists, such as the CS department’s grad list.

III. Headlines

NodeXL provides an implementation of the Clauset-Newman-Moore algorithm, which uses edge-betweenness to partition large networks into communities [2]. I used this to visualize the communities within my email correspondences:

![Email Communities discovered with CNM-clustering](image)

**Headline 1: UMD School Of Music divided**

Clustering produces a very accurate representation of the communities I am a part of. One interesting observation is that while most of these communities are cohesive, the UMD School of Music occupies two separate clusters, with one link between them. Both clusters include students and administrators from the School of Music. The separation of these clusters suggests some kind of social divide within this department. From manual inspection, the division seems to be along the lines of age in the case of students and sub-department in the case of faculty.
This past summer, I had the unique opportunity to play professional Ultimate (Frisbee) in the American Ultimate Disc League (AUDL). If we apply the Sugiyama layout to the “AUDL” group in Fig. 2, we can see an organizational hierarchy inherent in the emails sent out from players, coaches, captains, and owners:

![Fig. 3: Sugiyama Layout of AUDL group](image)

Fig. 3 shows captains in red, coaches in green, players in blue, and management in black. Nodes’ sizes correspond with their betweenness centrality. One can see that the majority of communication comes from the captains, sometimes first passing through management. The coaches occupy a position off to the side, with management. This is consistent with the team’s player-oriented approach.

![Fig. 4: Communication structure for two classes](image)
Fig. 4 shows emails sent between myself, instructors and other teaching assistants for two undergraduate classes. The visualization suggests two different organizational structures inherent in the two classes. In *Class 1*, the instructor is the center of communication, and information flows outward to the TAs. In *Class 2*, the organizational structure is “flatter;” everybody involved in organizing the class talks to everyone else, and the nodes form a clique.

IV. NodeXL

First of all, it was very interesting to see my email displayed as a graph, and once NodeXL was working it was very easy to set this up. NodeXL does a good job of explaining operations and filters that one can apply to the graph. The side-by-side display of the graph with the Excel cells makes it simple to understand what some part of the graph corresponds to in the data. The zooming also worked well, but it would be nice to also be able to scroll around once zoomed in.

NodeXL’s biggest shortcoming was its lack of portability. With only a Mac laptop available, getting NodeXL to work took between six and eight hours. I also encountered some difficult-to-interpret error messages while attempting to import emails. Additionally, if I did not significantly limit the size of the emails being imported, the import seemed to freeze indeterminately.

All in all, I enjoyed using NodeXL once I had set it up. Other possibilities for adding useful functionality might include:

- Allowing a user to see how a graph changes over time
- Making it so that double-clicking on a collapsed group expands that group
- Allowing the user to choose some of the clustering parameters
- Making sure that the names of filters are not cut off

As a side note, MIT’s “Immersion” [3] allows you to quickly make a graph of your Gmail contacts. Nodes (contacts) are sized by frequency of correspondence and clustered when they also correspond with each other.

V. References:

