CS171 Project Proposal

Background and Motivation

The Harvard SEAS website has a goal to provide visitors with an understanding of the inner workings of the school and the programs it offers. This involves content relating to faculty, degree programs, centers, initiatives, and research. In a typical school, these things are almost all invariably related to a departmental structure. SEAS is a little... different.

Our problem is that it's difficult for outsiders (like prospective students) to understand how SEAS works due to the cross-disciplinary nature of its degrees, classes, and research. We want to be able to show *connections* -- how the AP degree connects with studying robots that self-assemble or what biological phenomenons have to do with applied mathematics. We'd like to make these connections with teaching areas (our version of departments), research areas, degrees (both undergrad and graduate), and faculty members in such a way that allows for exploration of our offerings, entry from any of the four categories, and understanding of our structure that flat web content just doesn't allow.

As web dev for SEAS Communications, this falls under my area of expertise. Since the typical web structure wasn't working, a visualization seemed like the way to go.

Project Objectives

While we aren't trying to learn anything from our data, SEAS Communications want to answer common Admissions questions such as:

- What degrees are offered?
- Do my interests connect with your degree offerings?
- What faculty can I expect to work with while at SEAS?
- Who will I be able to do research with if I'm interested in X?
- How do your degrees relate to the types of courses offered (teaching areas)?

We hope that by visualizing the connections between these concepts it will be easier to understand than a web tree or a tabular presentation.

Ideally we'd like to trim the questions that come through Student Affairs and help incoming students better understand our somewhat unusual internal structure before they arrive.

Data

Our data comes mostly from tagging appropriate content types on our public-facing website (www.seas.harvard.edu). While not the case at the moment, future data will actually stem from a centralized data warehouse, but much of it now is manually maintained within the web CMS.   Some details (faculty directory information primarily) originate in Active Directory, but I'll be drawing everything from the website to simplify data collection.

Data Processing

I have already begun data cleanup (as well as some data creation – we did not yet have connections between programs and the other categories of content, for example). Data cleanup is part of the larger data warehouse project as well, so it's not only being done for the purposes of this project and therefore will be done at the source, rather than in post-processing.

Data will be collected and associated through a Drupal module (the code of which will be included in the final submission) and output as JSON for use in the visualization. The plan is to do heavy-duty processing through the module so that the JSON itself is as easy to use with the D3 libraries as possible. This should also speed display, as the JSON file can be cached.

Visualization

The expected visualization style will be a radial network graph. Some initial toying around with force diagrams and a squarish network graph proved less than optimal (more details on those experiments will come in the Process Book).
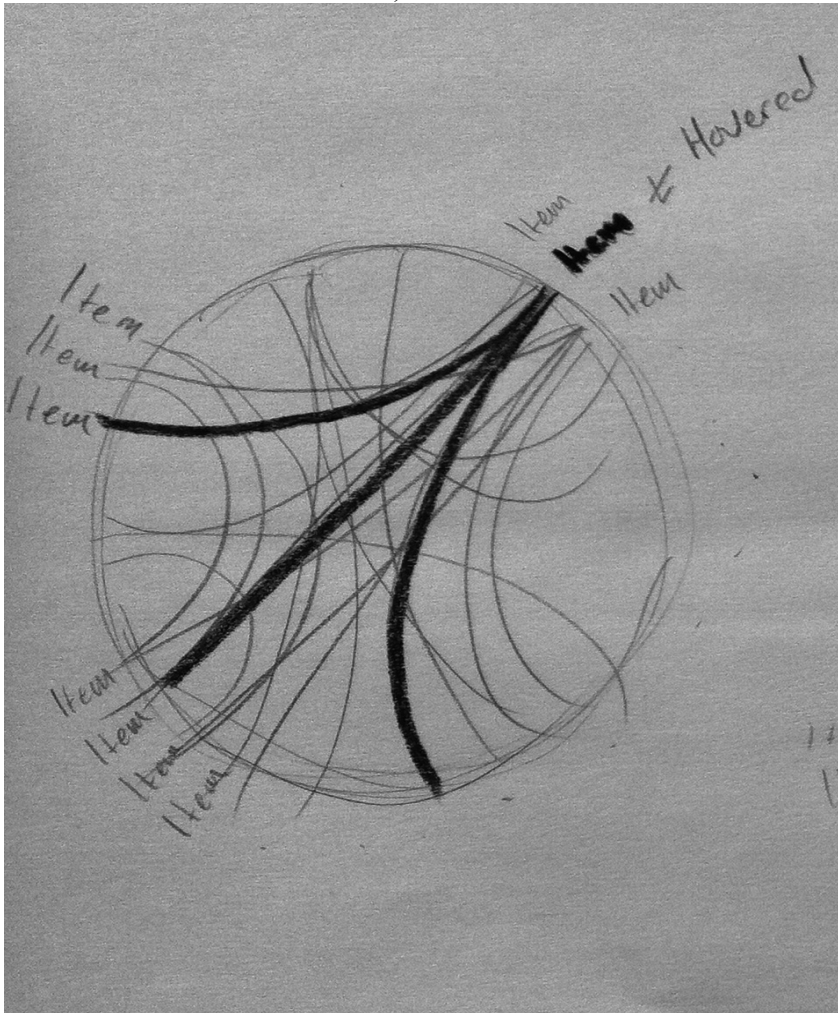
A detail view will be included. Exactly what this will look like is up in the air, but the prevailing design at the moment is to have the selected not come to center, grow, and include some metadata on

the node (eg. Faculty photo, category icon, a summary or details); additionally, the immediate connections to the selected node would move in and grow to emphasize them (and a click on one of them would roll it to center instead).
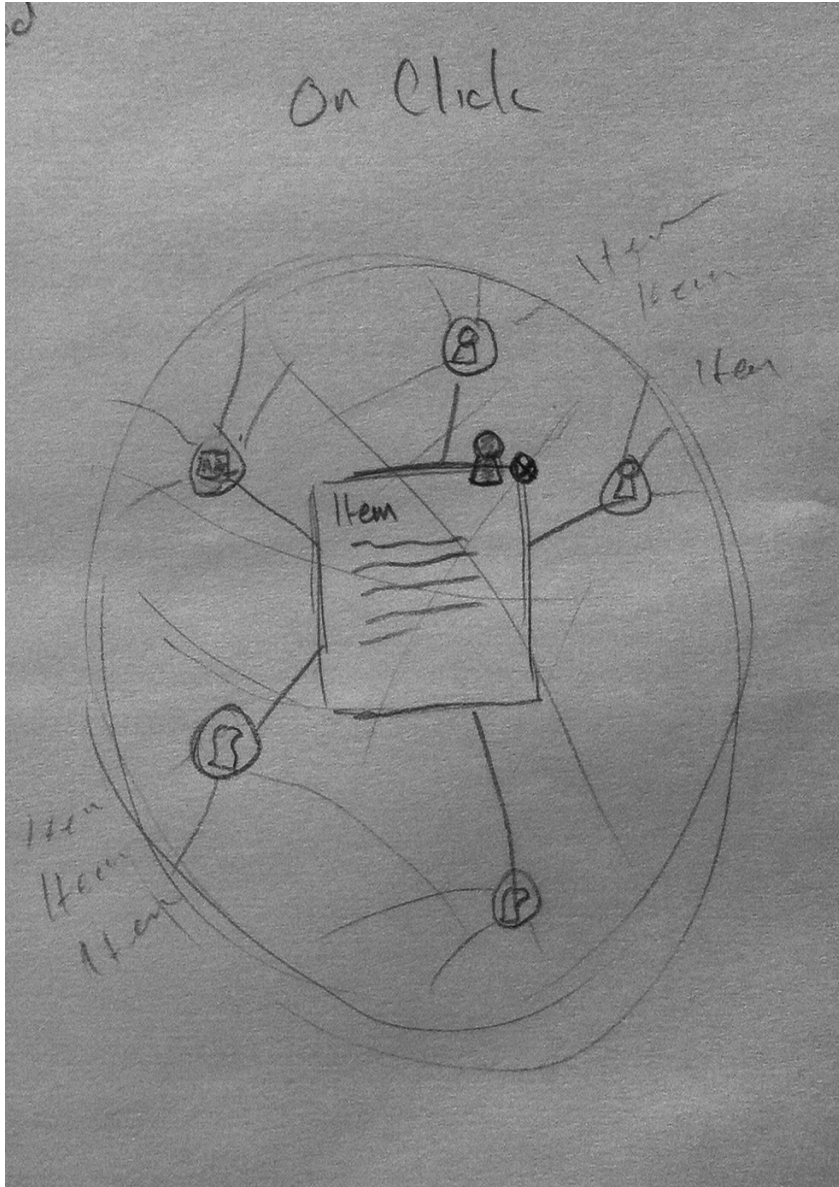
Must-Have Features

- **The Connections:** without these, there's no project; it is simply the most important aspect of this project.
- **Clear categorization:** you should be able to tell instantly what type of content you're looking at at any time (Faculty, Program, Teaching Area, Interest)
- **Tooltip/Detail View**: the name of the content is only so useful; some additional detail is crucial to answering the questions a prospective student will bring to the visualization
- **The ability to access full content detail:** Students much be able to jump off into the deeper web content if interested.

Optional Features

- **In-page search:** As-you-type highlighting of nodes to help attract attention to a particular piece of content. Particularly useful when trying to pinpoint a particular person or research interest.
- **Filtering**: Limit the graph so as to show only connections between particular categories. Useful if you have no interest in a given aspect
- **Centers and Initiatives:** a 5th aspect discussed by our group for possible inclusion; this data doesn't exist yet and might be overkill for the default view, but with filtering, we might be able to allow it to be selected in by choice.
- **Alumni outcomes/Careers:** a 6th aspect, relegated to a possible future feature for much the same reasons as Centers and Initiatives.

Project Schedule

Week 5	Finalize Data, determine what's missing
Week 6	Finalize Proposal/Sketches
Week 7	Proposal polishing/Proposal Due
Week 8	Check data in layout example code/Finalize detail view design
Week 9	Code detail view design
Week 10	Beautification/code refinement
Week 11	Prototype due
Week 12	Prototype review by TF and "clients"
Week 13	Revisions
Week 14	Final due