

RDF Graph Visualize Report

Everything is a graph

Think of the triple store as a graph. Slices of the triples store are sub graphs. Publication citation is a graph. A collection of triples with a given subject is a graph. A user profile is a graph.

R handles graph I/O very well

Poster from last year developed basic RDF crawling techniques. These are easily built out to crawl various entities and their attributes, yielding rich data structures that are easily processed by subsequent functions.

Specify domain -> Crawl the RDF to build an R stat net object -> Visualize or tabulate the stat net object.

No data will be stored in R. Everything is transient. No user interface work will be done. The R interface will be function calls of the form (The plot function in R is sub-classed, and can already draw stat net objects, the vivo2statnet function will return a stat net object):

```
g<-vivo2statnet(data-spec)
plot(g,plot-spec)
report(g,report-spec)
```

report will be a new function for representing graph data as a report with sort order, breaks, aggregation, titles, etc. Output formats will include XML (XSL-FO is possible for fancy formatting, but a future project), text, PDF (just a PDF of the text report, nothing fancy), CSV. CSV goal is reasonable ingest to Excel for hand finishing, if desired.

Statnet is a library of graph functions

Statnet provides graph primitives.

- Select data for the graph. Follow Ying's example of the sparql query generator. Add functions regarding collections of graphs ? Or is this just a super graph?
- Use statnet functions for graph manipulation.
- Use statnet functions for display
- Develop additional statnet functions for vivo processing and reporting – that is, the results of this work should be given back to the stat net project as libraries for inclusion with stat net downloads.

Report is a Graph is a Visualization

Create symmetry in these representations. Data is a graph. Show graph as a report. Show graph as a visualization. Report and graph being executed on a single underlying stat net data object. How far does stat net already go with this? A report is just a visualization of a graph? Common reports are visualizations of what graphs. Go back and forth between graph, viz and report.

Provide standard reporting functionality

Can reporting concepts be represented as attributes of graphs? Probably yes.

- Rows and columns
- Ordering
- Breaks
- Sorting
- Horizontal and vertical functions (use R apply methods so that all functions are available)
- Work from Graphs. No other data object? Likely some helper functions, but nothing at the top level.
- Export to Excel (find existing R function for this)

Examples of Use

- CTSA annual Use Cases
- College level reports of scholarship
- Organizational level reporting
- Personal reporting
- Reporting and visualizing the VIVO triple store.
- Facet reporting – pubs, grants, studies, concepts, etc.

What are interesting visualizations? What reports correspond? If we have a report, we can push a visualization button? Always?

Web accessible

The work should be done from a web interface. All examples should come from UF VIVO (if there are data problems we can fix them). But then point at Cornell, Ponce, etc.

A user interface is needed to select and processes, tabulate and draw. This could be a subsequent piece of the work.

This is mostly controller work. VIVO is the model. Views are static R outputs.

Once the controller is understood, the UI work will go faster – backwards, I know. Normally use cases come first, then data, then interface, then controller. In this case, we have data and the opportunity for an 80/20 set of functionality. The R work proposed here is opportunistic.

Position the work for publication

- Wherever statnet was published. Extending Statnet for VIVO analysis, for social network audience.
- Graph, Report, Viz. The R journal, for R advocates.
- For presentation at the VIVO conference.
- For presentation at informatics conferences – visualizing VIVO, VIVO for CTSA reporting,
- R functions for semantic web summaries. For a semantic web audience. How much of this work have they already done?

Notes

1. Name of the project is ridiculous. Please suggest something cool.