

A tutorial on (my) research codes for graph (and sparse matrix) informatics

David F. Gleich

Stanford University

*College of Charleston
Clustering and Ranking Workshop*

August 15th, 2009

Funded by Microsoft Live Labs Fellowship
and the Library of Congress

Overview

MatlabBGL “Boost graph library” in Matlab

libbvg bvgraphs in C and Matlab

gaimc graph algorithms in pure Matlab

vismatrix super-spy (*older version public*)

gbrowse look at a graph (*not public yet*)

rapr random-alpha PageRank

innout inner-outer iterations for PageRank

netalign network alignment

Packages

Publications

boring

Research software
for informatics

catchy

A sparse matrix trick a day
keeps the problems at bay

trendy

Making your own tipping points
with small software tools

Philosophy

Top down

Fully conceived idea

Easy to use
(Powerpoint/Keynote)

Grow into monsters

Bottom up

Collects of programs

Sometimes difficult to use
(unix command line tools?)

Interdependence and duplication

Matlab?

Types of problems

Large, sparse data

Matrix or graph

Transposable data

Dydactic data

(not tensors, sorry
Mike, Tammy, Lek, etc.)

MatlabBGL

Bring the Boost graph library to Matlab

A square sparse matrix is the adjacency matrix of a graph

No copies of data when possible
(or when forced by the user)

```
A = grid_graph(6,5);  
xy = kamada_kawai_spring_layout(A);  
gplot(A,xy);
```

```
A = erdos_reyni(100,0.05);  
[cc sizes] = components(A);  
max(sizes)
```

libbvg

Web graphs are highly compressible

Work with these things in C instead of Java

And Matlab too. (Anyone want to help write a Python wrapper????)

```
G = bvgraph( 'webbase-2001' ); % 10M-by-10M network  
d = sum(G,2);
```

```
A = sparse(G); % oops!
```

```
X = bvpagerank(G,0.85); % what else?
```

gaimc

MatlabBGL is big, hard to compile, and hard to change

Graph algorithms in Matlab code

“But for loops in Matlab are slow!”

Maybe not too slow... (2-4x? 3-8x?)

```
A = load_gaimc_graph('celegans');  
cc = dirclustercoeffs(A); % directed clustering coeffs
```

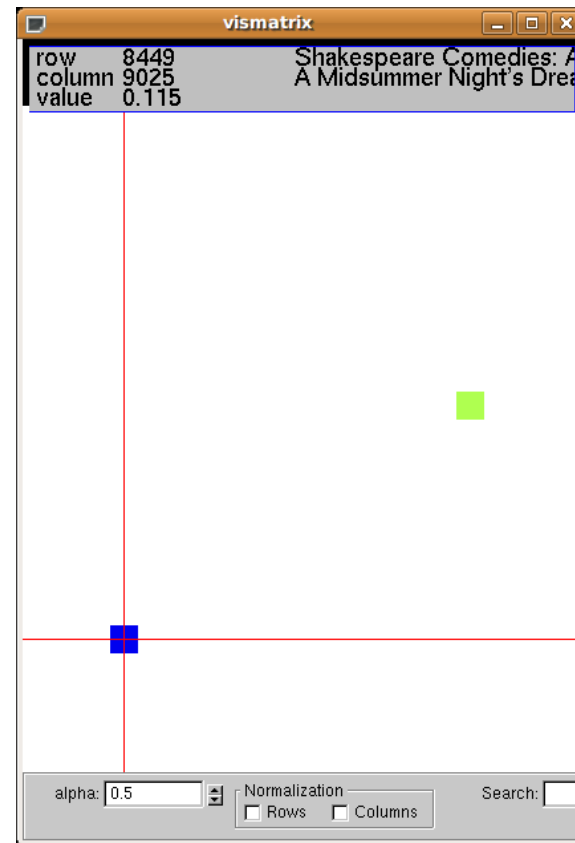
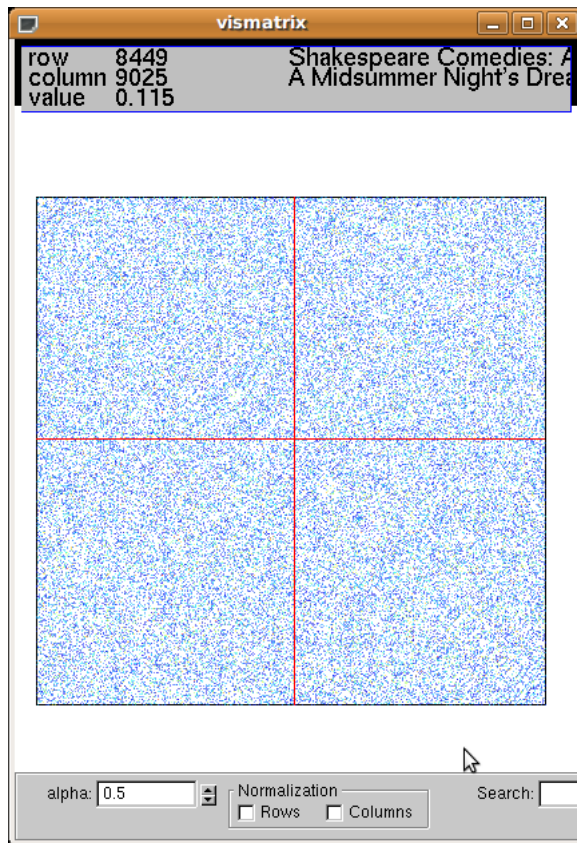
```
A = load_gaimc_graph('airports');  
cn = corenums(A);
```

vismatrix

OpenGL based sparse matrix visualization

Labeled rows and columns

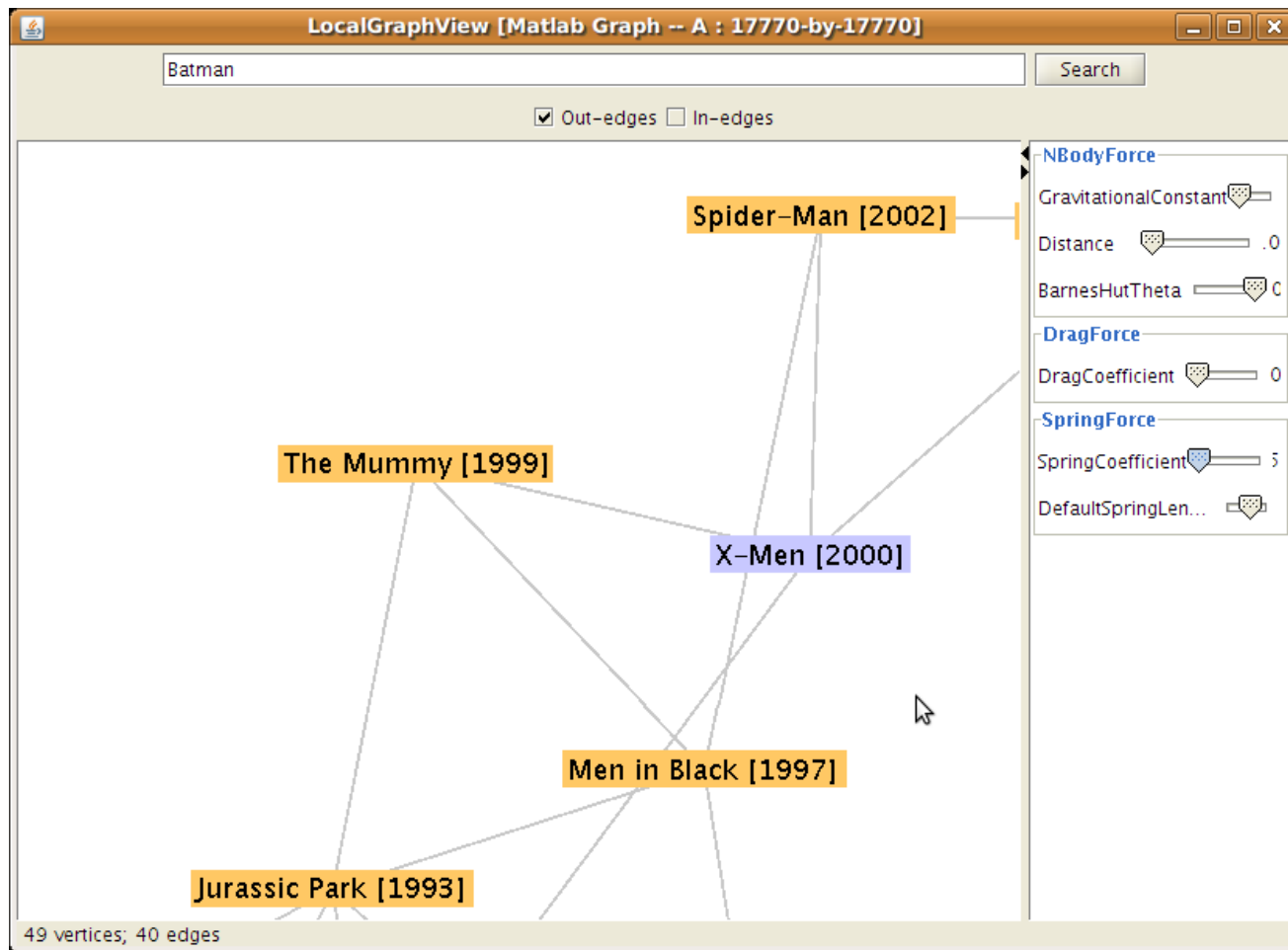
Interactive browsing



gbrowse

Just look at a labeled graph

Uses prefuse library of interaction



Publication software

For each recent paper, all the experimental code is public.

YOU shouldn't have to take *my* word for *my* results

```
$ ls
```

```
data/  
experiments/  
matlab/  
README.txt
```