New ideas for network analysis

Bryan W Lewis, R/Finance 2014

http://goo.gl/ZbJEUO
The work of:

Baglama, Benzi, Boito, Estrada, Fenu, Golub, Kleinberg, Meurant, Reichel, Rodriguez, and others...
A simple undirected network
Represented as an adjacency matrix $A$
\[ [A^m]_{i,j} \]

Counts the number of paths of length m between nodes i and j
\[ [A + A^2 + A^3 + \ldots]_{i,j} \]

Counts paths of all lengths between nodes \(i\) and \(j\)

(may be unbounded)
Maybe long paths are not as important as short ones?
De-emphasize long paths by weighting them
\[ w_0 I + w_1 A + w_2 A^2 + w_3 A^3 + \ldots \]
When \( w_j = 1/j! \), then

\[
\begin{aligned}
w_0 I + w_1 A + w_2 A^2 + w_3 A^3 + \ldots &= \exp(A) \\
\end{aligned}
\]
Some useful measurements of networks

for functions $f$ that are nonnegative, nondecreasing on the spectrum of a symmetric matrix $A$, for example $f(A) = \exp(A)$
Importance of a node in the network

f-subgraph centrality

$[f(A)]_{i,i}$
Ease of travel between two nodes

$f$-subgraph communicability

$[f(A)]_{i,j}$
Starting nodes giving greatest network coverage

$e_i^T f(A) c$
Directed networks ➔ Nonsymmetric matrices
In that case we can work with

\[
\begin{bmatrix}
0 & A \\
A^T & 0
\end{bmatrix}
\]

Related to the HITS hub/authority approach
In that case, we compute

- $f$-subgraph hub centrality
- $f$-subgraph authority centrality
- $f$-subgraph hub communicability
- $f$-subgraph authority communicability
- ...
Computing this stuff
When \( f(A) = \exp(A) \) we can use `expm` from the R's superb Matrix package.
Benzi and Boito:

Use orthogonal polys from quadrature rules (Golub)
What if we only care about the top $m$ important nodes?
IRL methods project into a subspace guaranteed to contain the most important nodes.
Demo
bitcoin
transaction
network (directed)

Ivan Brugere
(UIC!!)

http://compbio.cs.uic.edu/data/bitcoin/
6.3 M x 6.3 M adjacency matrix

16.3 M nodes
Compute top 5 hub centralities

```r
> load("bitcoin_from_to_graph.rdata")
> t1 <- proc.time()
> x  <- topm(B,q=2,tol=0.1,m_b=5)

> proc.time() - t1
  user  system elapsed
  86.970  24.350 111.605
```
Small problem comparison
Leading 1000 x 1000 submatrix of bitcoin data

```r
> t1 <- proc.time()
> ex <- diag(expm(X) + expm(-X))/2
> proc.time() - t1

user  system elapsed
151.080   0.220 151.552

> i <- order(ex, decreasing=TRUE)
> i[1:5]

[1] 11 25 27 29 74

> t1 <- proc.time()
> top <- topm(X, type="cent")
> proc.time() - t1

user  system elapsed
0.555   0.010   0.565

> top$hubs

[1] 11 25 27 29 74
```
Plan

New IRL package will contain at least

- IRLBA (truncated SVD)
- IRBLB (windowed truncated SVD)
- IRBLeigs (symmetric variant)*
- toppm (network bounds)
- others...

Succeeds current IRLBA package
Some new research in process...
https://github.com/bwlewis/IRL

These slides are available here:

http://goo.gl/ZbJEUO