

Approximated Parallel Betweenness Centrality

Geoffrey Foster

School of Computer Science
Carleton University

November 22, 2007

Social Network Analysis

- Related to Network Theory
- A technique in sociology, anthropology, information science, etc.
- Small world phenomenon
 - 6-degrees of separation (1967-Stanley Milgram)

What can be measured?

- Centrality
- Structure of Communities
- Categorization/Grouping
- etc.

Why bother?

- Marketing research
- Counter-terrorist and Intelligence operations
- Analyze the spread of disease/infection
- For Fun?

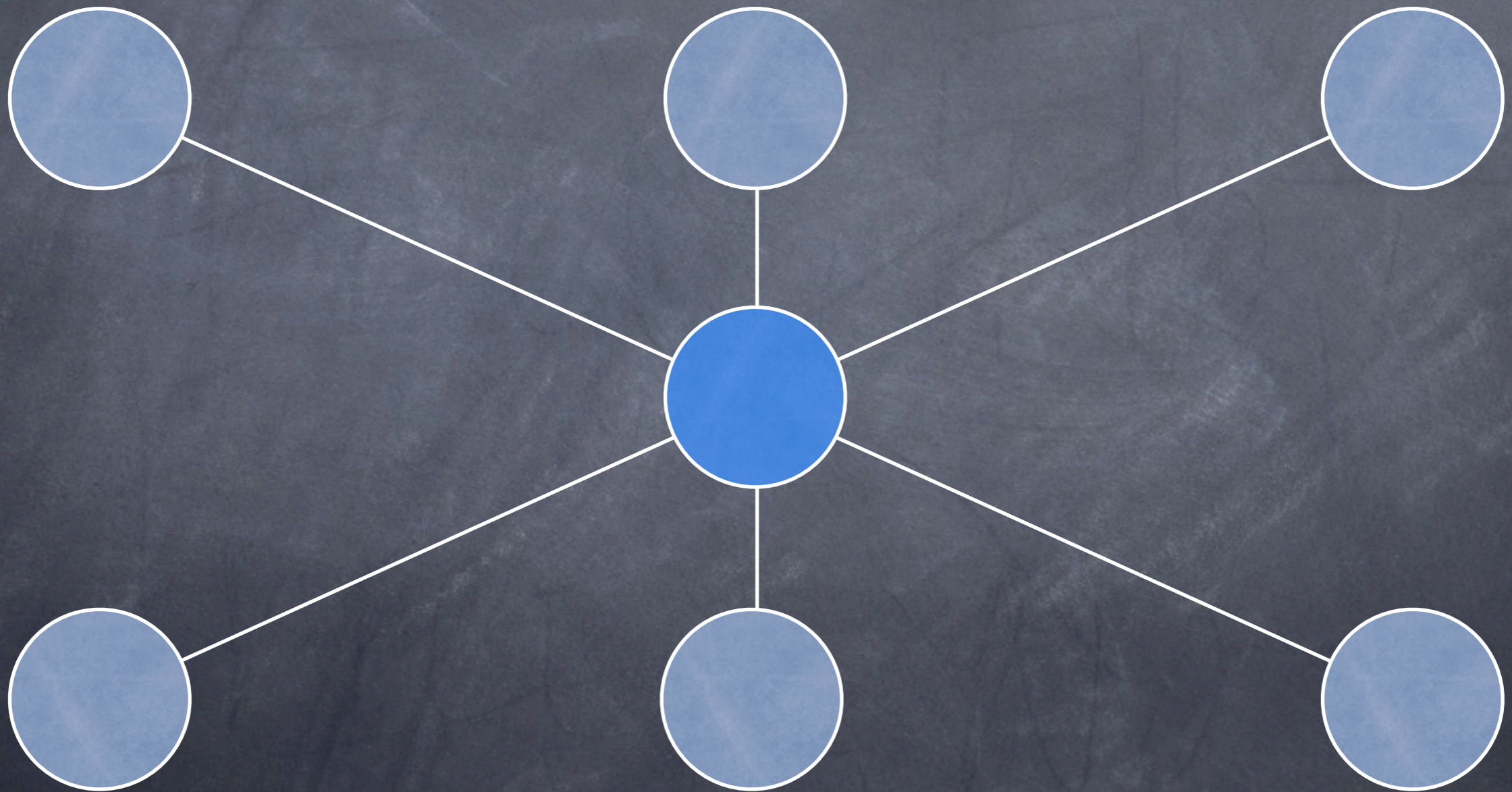
What is Betweenness Centrality?

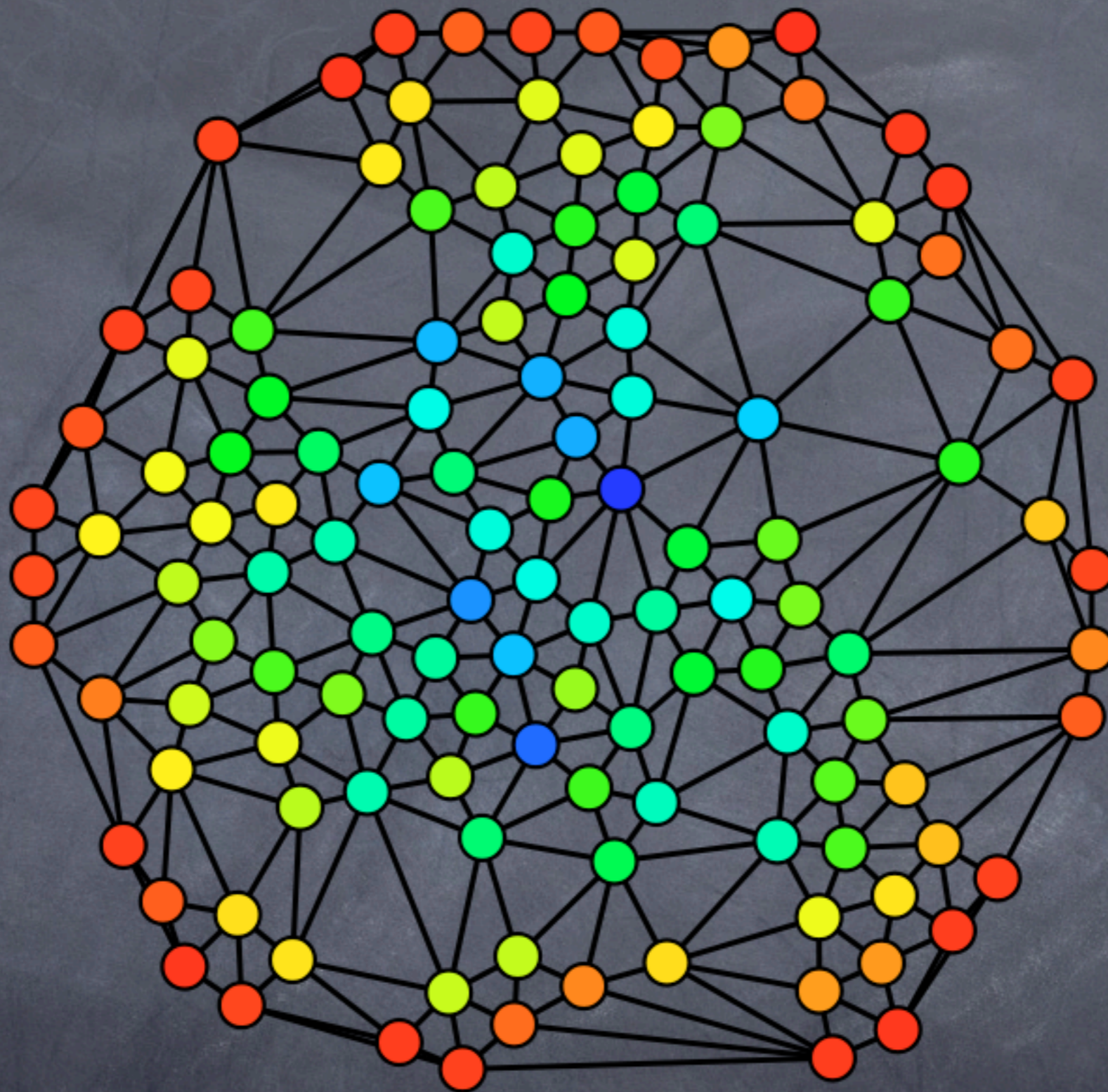
- “node with high betweenness has great influence over what flows -- and does not -- in the network.” (<http://www.orgnet.com/sna.html>)
- A measure of how “important” a vertex in a graph is

In Mathy terms...

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Example





Hue (from red=0 to blue=max) shows the node betweenness.

What is it useful for?

- Identifying:

- bottlenecks in networks

- central figures in organizations

- terrorist or criminal

- Marketing

- Disseminating information to a group

Sequential Algorithm

- For every vertex $v \in V$:
 - Perform a Breadth-First-Search with v as source
 - For any vertex, w , record the number of shortest paths that go through it
- Betweenness Centrality of v is given by the previously seen formula for $C_B(v)$

Sequential Algorithm Analysis

- Best known sequential algorithm
- $O(n + m)$ space
- $O(nm)$ run-time ← We can do better!

Real-World Data-set

- Sequential tested on randomly generated undirected, unweighted graph
- 100 to 2000 vertices of varying densities

= (That's not very realistic...

Solution

- “Borrow” a real social networks data
- Facebook! ← Estimated 55 Million users
- Obtained “friend” knowledge for roughly 3 Million users thus far

Parallel Betweenness Centrality

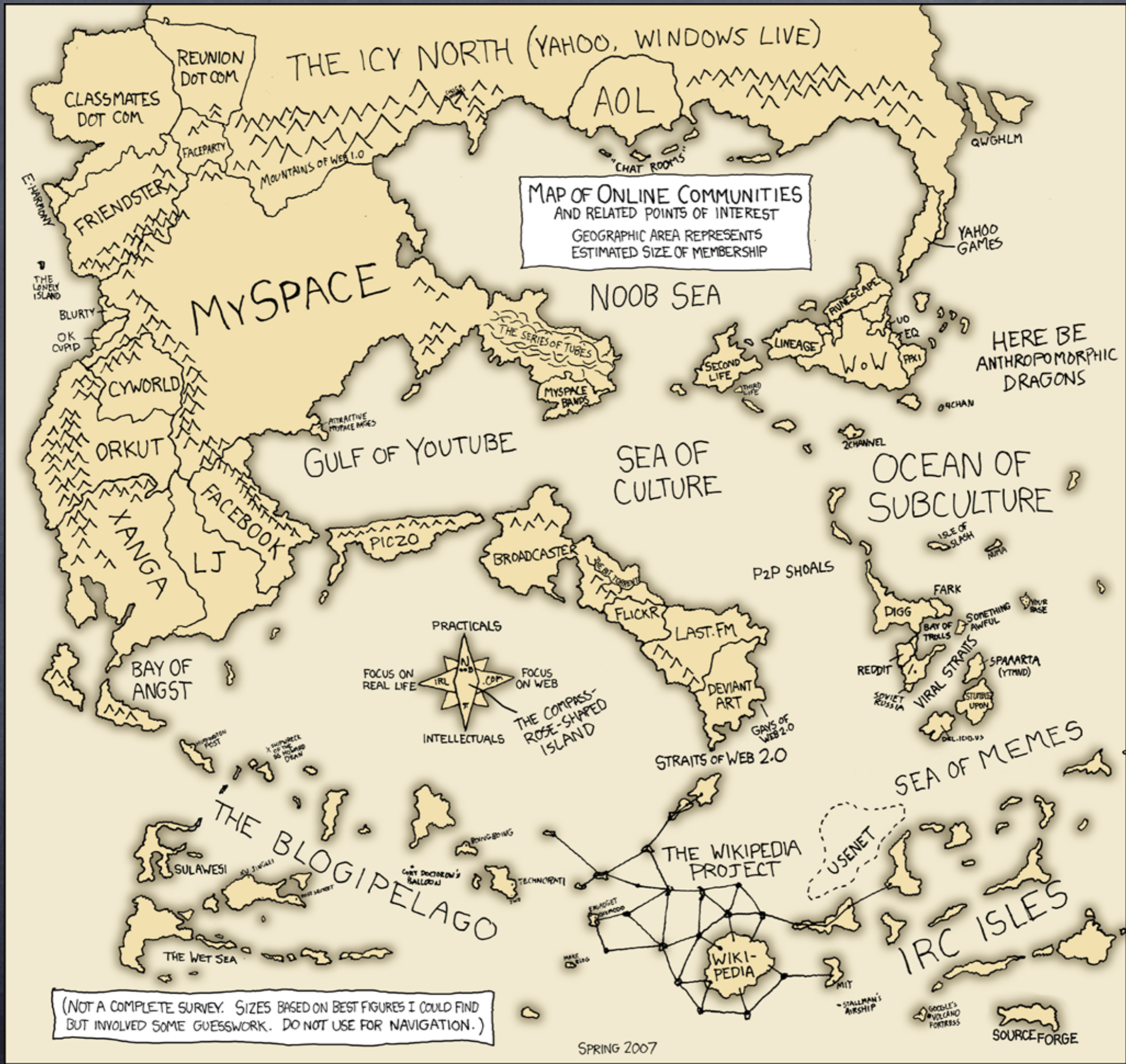
- Easily Parallelized to CGM model
- Same as Sequential, but:
 - $\frac{|V|}{p}$ vertices sent to each processor
 - Each processor does a BFS for its $\frac{|V|}{p}$ vertices
 - Results are combined

Parallel Approximated Betweenness Centrality

- Use Parallel version of Brandes' Algorithm
- Calculate for a subset of K vertices selected at random
- Perform approximation based upon computed values

$$\hat{C}_B(v) = \sum_{i=1}^K \frac{n}{K} \delta_{v'_i}(v) \quad \delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Questions?



MAP OF ONLINE COMMUNITIES
AND RELATED POINTS OF INTEREST
GEOGRAPHIC AREA REPRESENTS
ESTIMATED SIZE OF MEMBERSHIP

(NOT A COMPLETE SURVEY. SIZES BASED ON BEST FIGURES I COULD FIND
BUT INVOLVED SOME GUESSWORK. DO NOT USE FOR NAVIGATION.)

SPRING 2007