The Shape of the Web

So, the Web is a directed graph, but what does it look like?

What is the shape of the web?

Broder et al.: Graph structure of the web (2000)

The Bow-Tie Shape of the Web

- **CORE**: A giant strongly connected component (SCC)
- **IN**: The part of pages leading to CORE
- **OUT**: Pages reachable from CORE
- **TENDRILS**: Not part of CORE, leading out of IN, into OUT, or bypassing CORE (TUBES).
- **ISLANDS**: Disconnected pages.

How do you compute a SCC?

Within a SCC, a path can be found from any node to any other node: 
1 $\rightarrow$ 2, 2 $\rightarrow$ 3, 3 $\rightarrow$ 1.
**Strongly Connected Component**

- SCC is defined on a digraph $G = (V,E)$ only! (Why?)
- SCC is a subset $C$ of $V$ with the following properties:
  1. $\forall u,v \in C$, $u$ is reachable from $v$ in $G$ and $v$ is reachable from $u$ in $G$
  2. If $C$ is a proper subset of another subset $D$ of $V$, then $D$ does not satisfy property 1

**Translation:**
C is a maximal subset of Vertices with mutual reachability

Which graph traversal algorithm(s) produce reachability information?

**Transpose of a Digraph**

- If $G = (V,E)$ is a digraph, then its transpose $G^T$ is the digraph $G = (V, E')$ where $E' = \{(v,u) \mid (u,v) \in E\}$

**Which graph has more SCCs? $G$ or $G^T$?**

**A fact of life...**

- A directed graph and its transpose have exactly the same strongly connected components

**Why?**

**Detecting the STRONGLY CONNECTED COMPONENT requires 2 DFS traversals**

1. Run DFS to compute the finishing times $f(u)$.
2. Compute the graph’s transpose.
3. Run a 2nd DFS, while considering vertices in the order of decreasing $f(u)$.
4. Output the vertices in each tree in the forest as a separate SCC
SCC identified. How about the rest?

IN:
OUT:
TENDRILS:
ISLANDS:

Visualizing a small Web

But Why is it a Bowtie?

Hey is a teapot, a daisy?
A bugle? A cauliflower?

It is a collection of Bowties,
because.
(it could not be anything else)

Proof by construction

M: Why the Shape of the Web is a Bowtie? (2010)

Bowtie Web: Proof by Construction

Start by considering one link per page
Pseudo-trees appear
The cycles of pseudotrees are like budding COREs
INs created, no OUTs
The Second link creates a Bowtie

- Consider the 2nd link
- It will reduce the number of components,
enlarge the CORE,
create IN and TENDRILS
- OUTs may appear
  as smaller cycles (than CORE)

Nodes w/out links and Third links
- Now include nodes w/out links
  (as possible targets for Bowtie nodes).
- They start off as ISLANDS
- Consider the effect of the 3rd link.
- What happens when you link:
  - IN node to IN, CORE, OUT, ISLAND
  - CORE node to IN, CORE, OUT, ISLAND
  - OUT node to IN, CORE, OUT, ISLAND
  - ISLAND node to IN, CORE, OUT, ISLAND

Third links enlarge Bowties!

Web is many Bowties!