What to do with directed edges?

PageRank, personalized pageRank (ppr), and two-way analysis

Karl Rohe; 992

We often imagine graphs as undirected.

Directed Graphs have new demands

- Edges are no longer symmetric. This means two things.
- Centrality is easier
- “symmetry” is one touch harder

PageRank gives each node a measure of centrality

It is defined as the "stationary distribution of a Markov chain"

- Imagine each node is a website and each edge is a hyperlink. directed!
- In the random surfer model, you start at a page chosen uniformly at random. Then repeat this process:
    - With probability .85, click on a link chosen randomly on the webpage
    - With the other .15, "teleport" to a webpage chosen uniformly at random
- Imagine doing this 1,000,000 times. The pageRank value for page i is the probability that you land that page on the 1,000,000 click. (1M is not special here, just a big number)

PageRank can be computed very quickly

- Because it can be computed quickly, it scales to very large graphs.
- The founders of Google developed this algorithm and it was the first version of the google algorithm for ranking webpages.
- You would give a search term, then it would find a subgraph of pages that used your search terms, then compute pageRank on that subgraph.
    - (I don’t think that is true… it is just a story. It is also possible that they used something more like “personalized pageRank”)

Personalized pageRank

Teleport to a specific node every time

- when you teleport in standard pageRank, you select a node uniformly at random.
- In personalized pageRank (ppr), you always teleport back to a specific “seed node”
- Then, each node has a ppr vector, where that node is the seed node.
- What might you want one of these for?

Personalized pageRank can be computed with only “local information”

Approximate ppr (appr)

- Imagine the massive twitter graph
- We don’t have access to it all, but we can request the neighborhood of any node (who do they follow)
- There is a fast way to approximate the ppr vector for a node, where you only need to know the neighborhood information for a handful of nodes (bigger handful, better approximation)
- https://github.com/rohelab/appr
- This is how we sample twitter for the murmuration. **Targeted sampling**

What else to do with directed edges?

~~PageRank, personalized pageRank (ppr),~~ and two-way analysis

Karl Rohe; 992

What do communities mean?

- Are directed edges a notion of similarity?

Making a notion of similarity (symmetric!) from directed edges

- Common downstream-neighbors
- Common upstream-neighbors

- What do I mean by symmetry?

Row clusters and column clusters

- $E(A) = ZBY^T$
- Row clusters in Z (how nodes send edges)
- Column clusters in Y (how nodes receive edges)