The Role of Homophily and Popularity in Informed Decentralized Search

Denis Helic & Florian Geigl
Knowledge Technologies Institute
Infeldgasse 13/5. floor, 8010 Graz, Austria
{florian.geigl,dhelic}@tugraz.at
http://kti.tugraz.at/

September 15, 2014
The Role of Homophily and Popularity in Informed Decentralized Search

- Decentralized Search

- Informed Decentralized Search
  - steered by some kind of knowledge
The Role of *Homophily* and *Popularity* in Informed Decentralized Search

**Homophily**

- Start
- Target

**Popularity**

- Start
- Target

Denis Helic & Florian Geigl
September 15, 2014
Motivation

- large networks
- dynamic networks
- no central search
  - P2P
  - swarm of drones

Stackoverflow.com
Communication-Network
Related Work

Kleinberg

fixed mixture: Jensen

Adamic
Proxies

Homophily:

\[ 0 \leq \text{cosine similarity} \leq 1 \]

[common neighbours(i,j)]
\[ \sqrt{\text{degree}(i) \times \text{degree}(j)} \]

Popularity:

degree of the node
Example: Greedy Navigation using Popularity
Normalization

Start

\[ \frac{4}{12} \]

\[ \frac{8}{12} \]

Target

Denis Helic & Florian Geigl
September 15, 2014
Mixture Distribution

\[ \text{mixture} = p^*\alpha + q^*(1 - \alpha) \]
Mixture Distribution

![Graph showing mixture distribution with different alpha values: 
- P, alpha = 0.25
- H, alpha = 0.5
- alpha = 0.75](image)
Datasets

- DBLP
- Facebook Subset
- Twitter Subset
- Wikipedia for Schools

nodes $\sim 4k - \sim 300k$
random missions

vary $\alpha$ from 0 to 1

Success Rate
Results Greedy Navigation

mixture: $H^\alpha + P^{*(1-\alpha)}$
Background Knowledge Models

- static mixture ✔
- static switch
  - inspired by human navigation

Start → Step 1 → Step $x$ → Step $n-1$ → Target

$\alpha = \text{Initial } \alpha$

$\alpha = 1 - (\text{Initial } \alpha)$
Results Greedy Navigation
Background Knowledge Models

- dynamic switch

\[ \alpha = \text{Initial } \alpha \quad \alpha = 1 - (\text{Initial } \alpha) \]
Results Greedy Navigation

mixture: \( H^\alpha + P^{1-\alpha} \)
Navigation Models

• greedy search
  • always use best

• stochastic search
  • draw out of mixture distribution

• softmax search:
  • apply softmax on convex combination
  • draw out of resulting distribution
Softmax
Results Stochastic & Softmax

stochastic search

softmax search
Discussion

- Homophily seems to be more important
  - degree distribution
  - low diameter networks

- cosine similarity includes a lot of information
When searching your „node“, don‘t pick the popular ones, take the similar 😊