The Role of Homophily and Popularity in Informed Decentralized Search

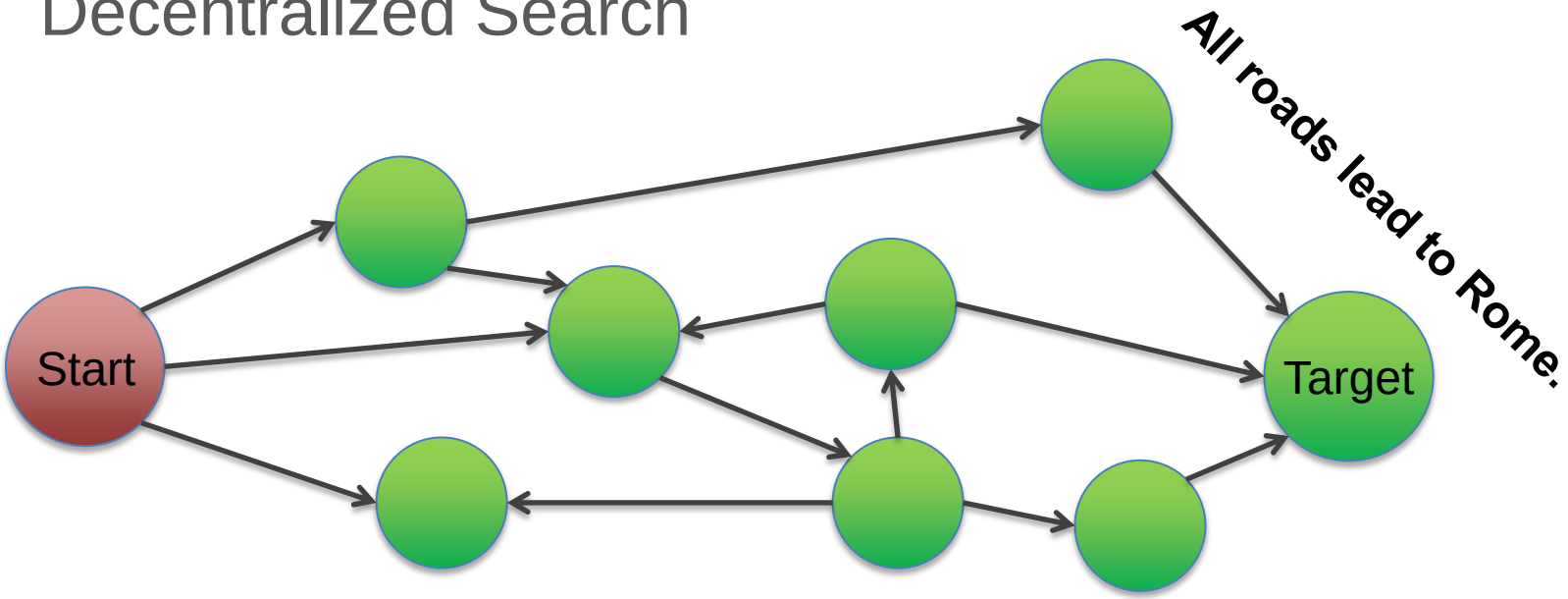
Denis Helic & Florian Geigl

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

September 15, 2014

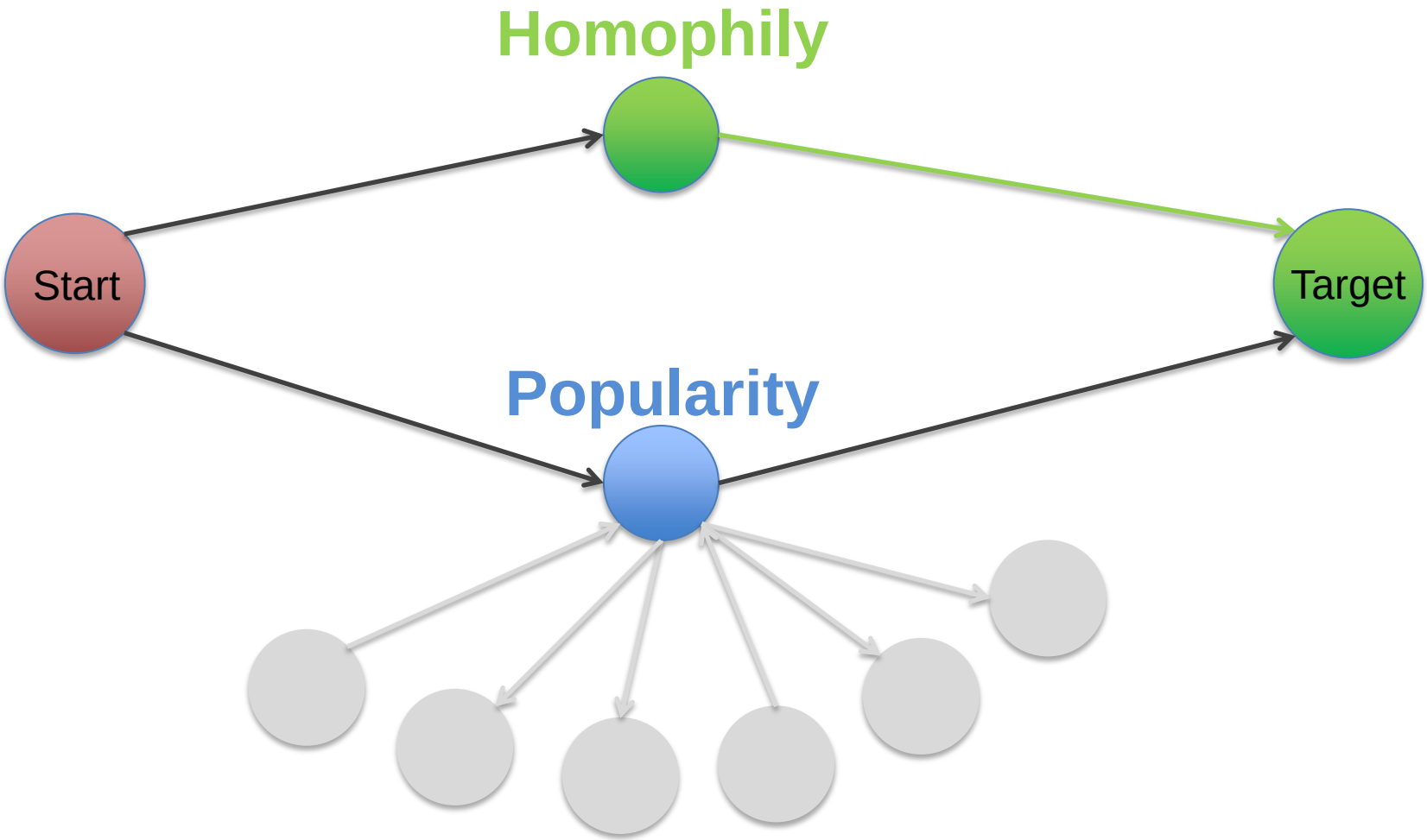
The Role of Homophily and Popularity in ***Informed Decentralized Search***

- Decentralized Search



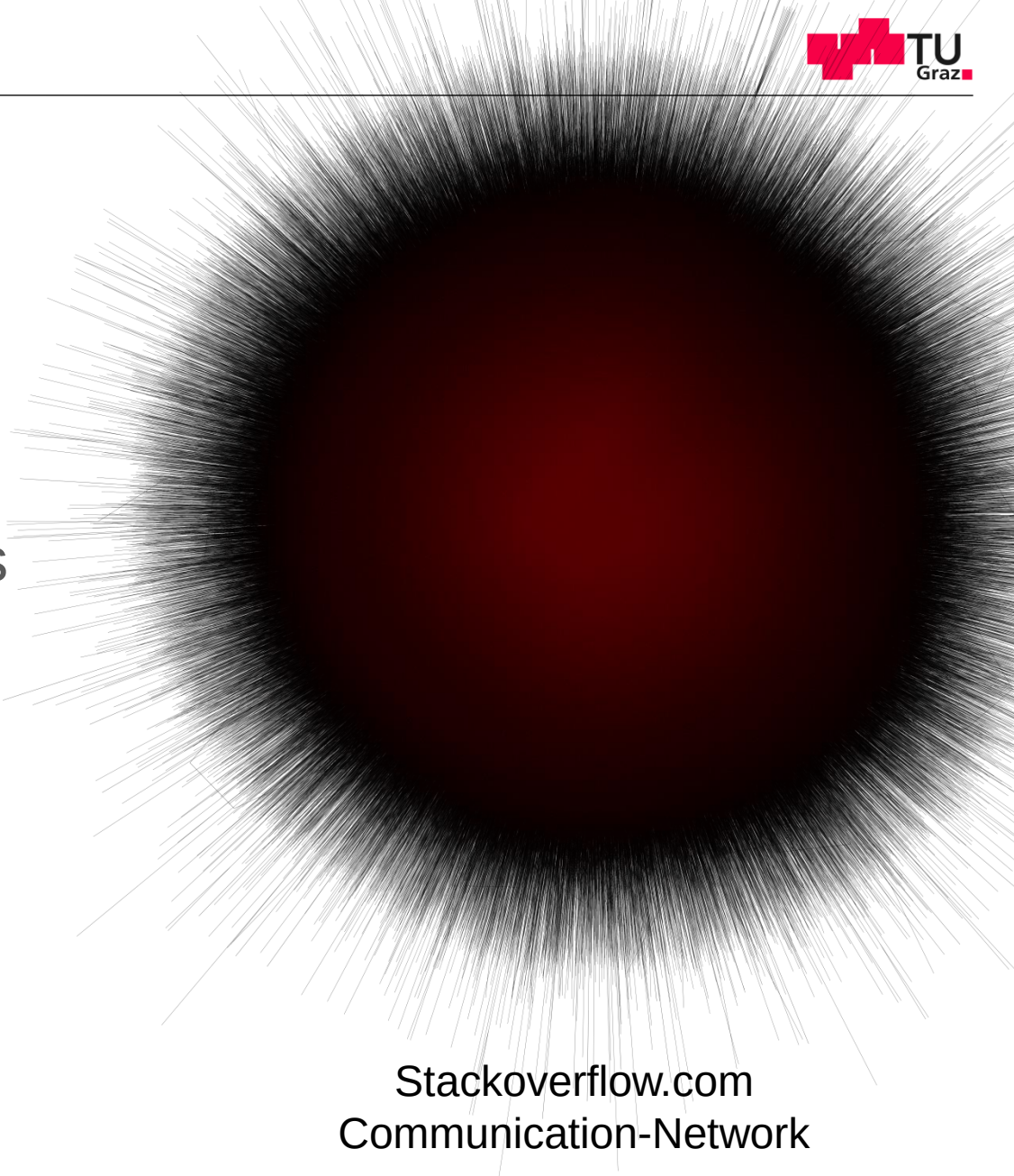
- Informed Decentralized Search
  - steered by some kind of knowledge

The Role of *Homophily* and **Popularity** in Informed Decentralized Search



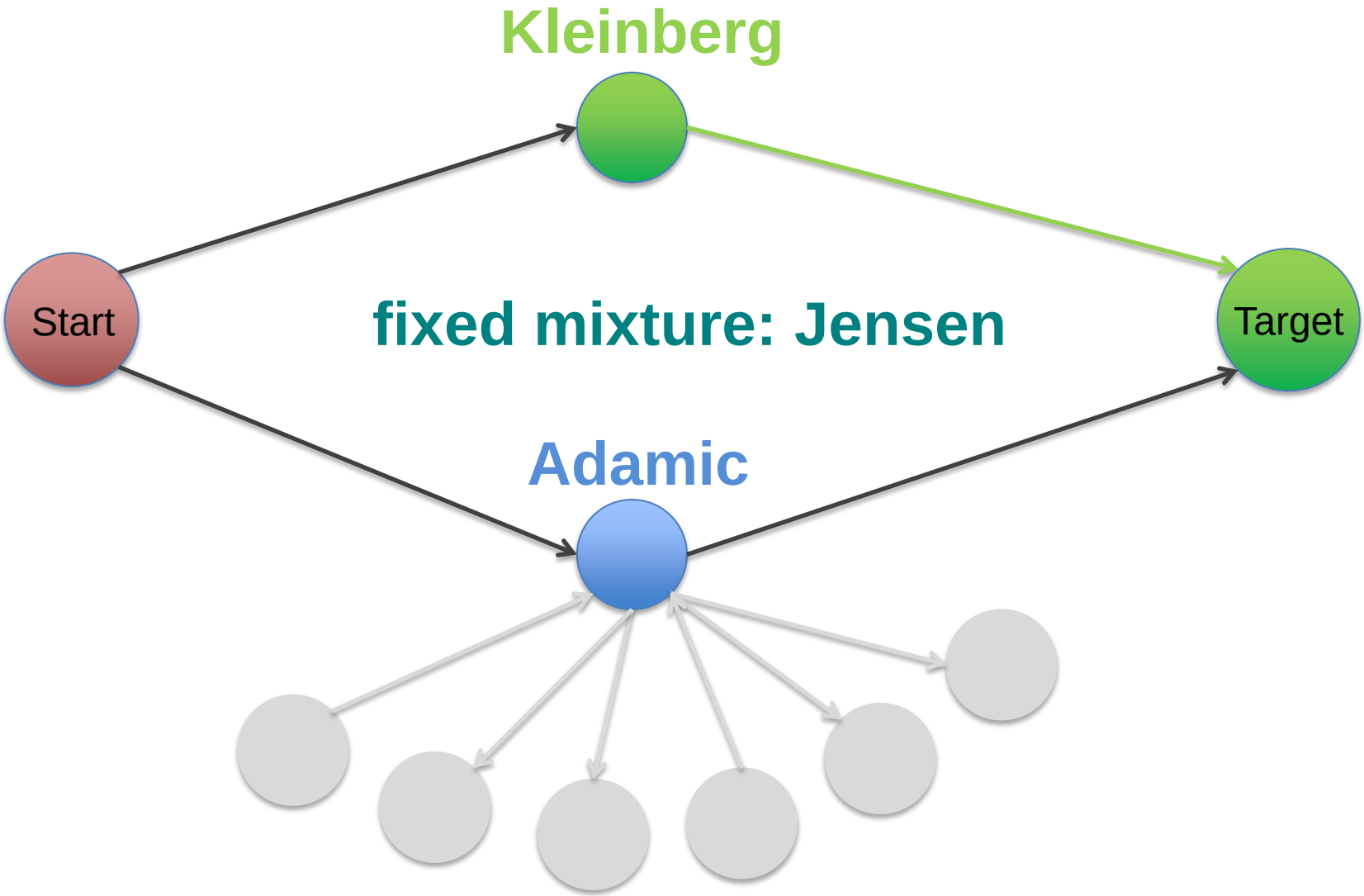
Motivation

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



Stackoverflow.com
Communication-Network

Related Work



Proxies

Homophily:

cosine similarity to target node

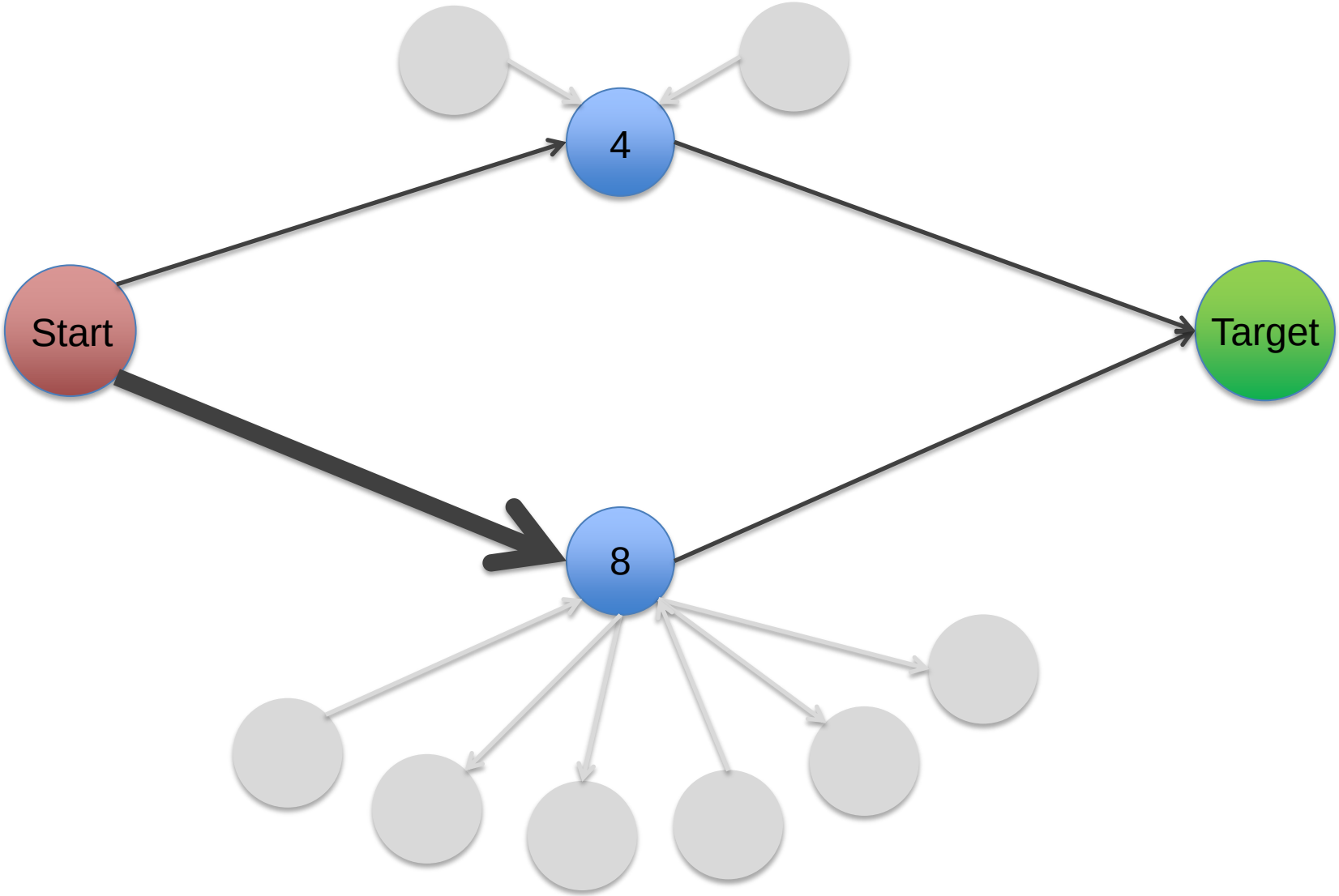
0 <= cosine similarity <= 1

$$\frac{common\ neighbours(i,j)}{\sqrt{degree(i)*degree(j)}}$$

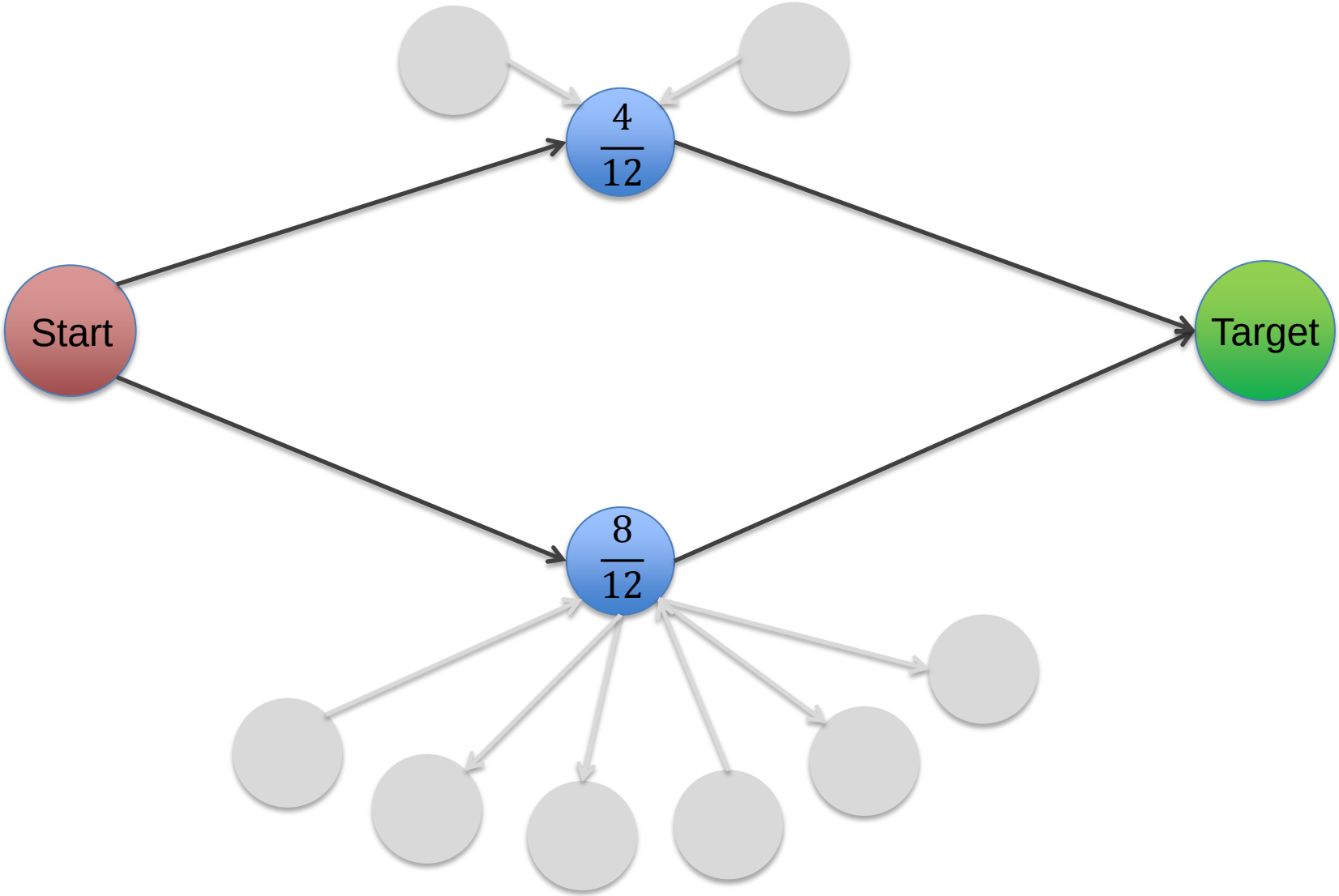
Popularity:

degree of the node

Example: Greedy Navigation using Popularity

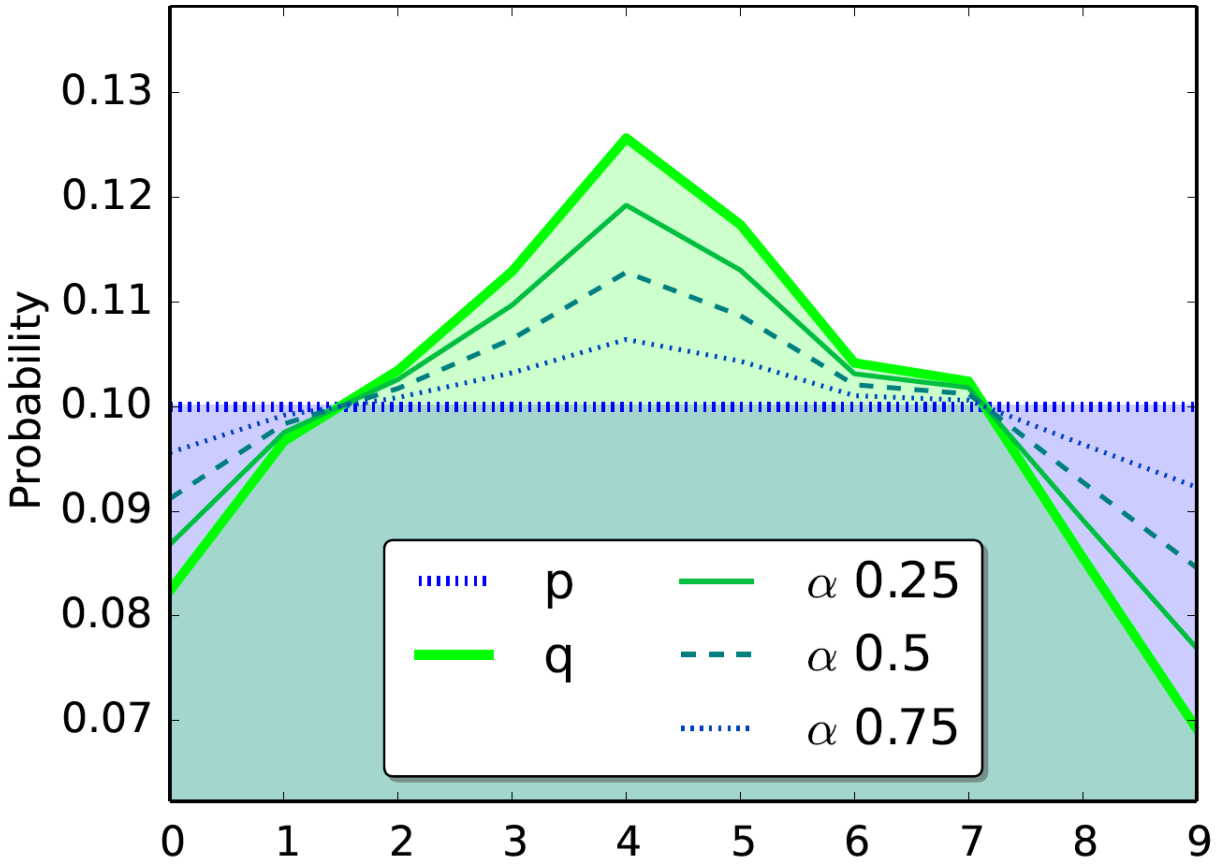


Normalization

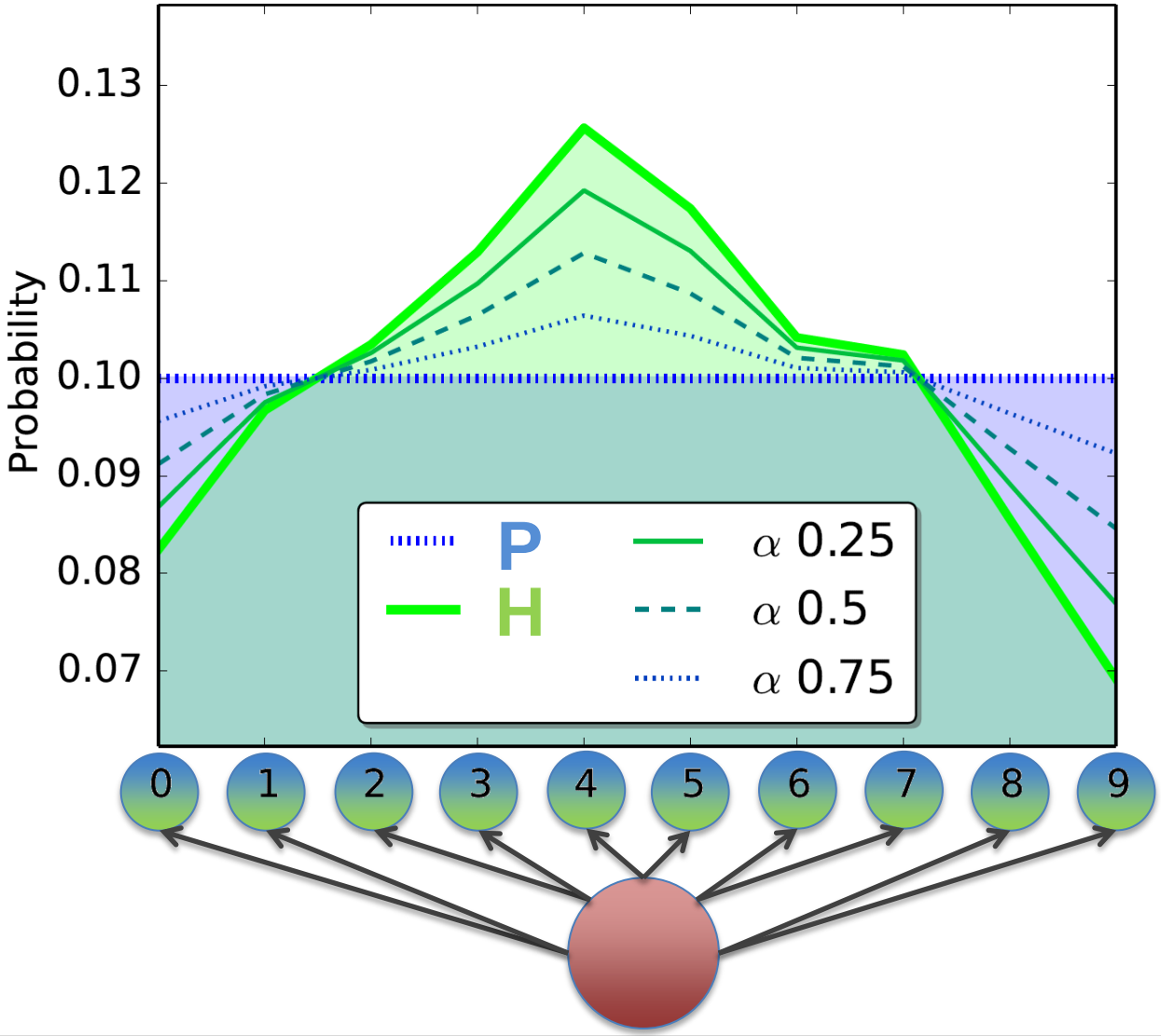


Mixture Distribution

mixture = p*α + q*(1- α)

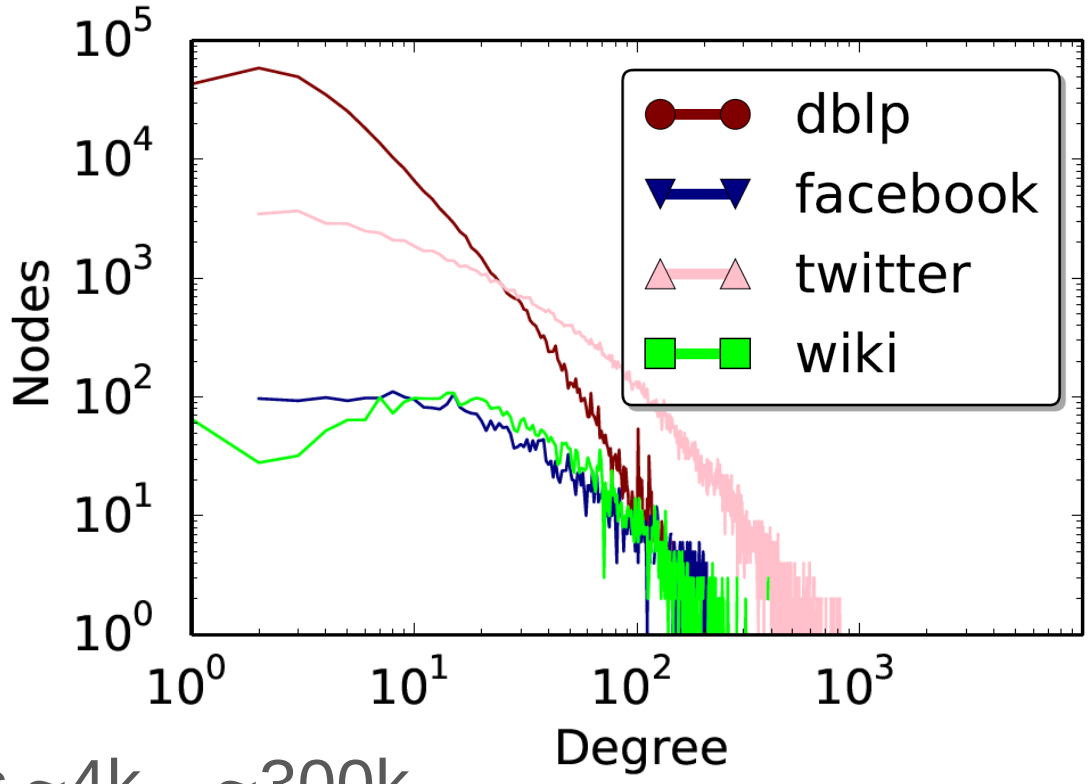


Mixture Distribution



Datasets

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



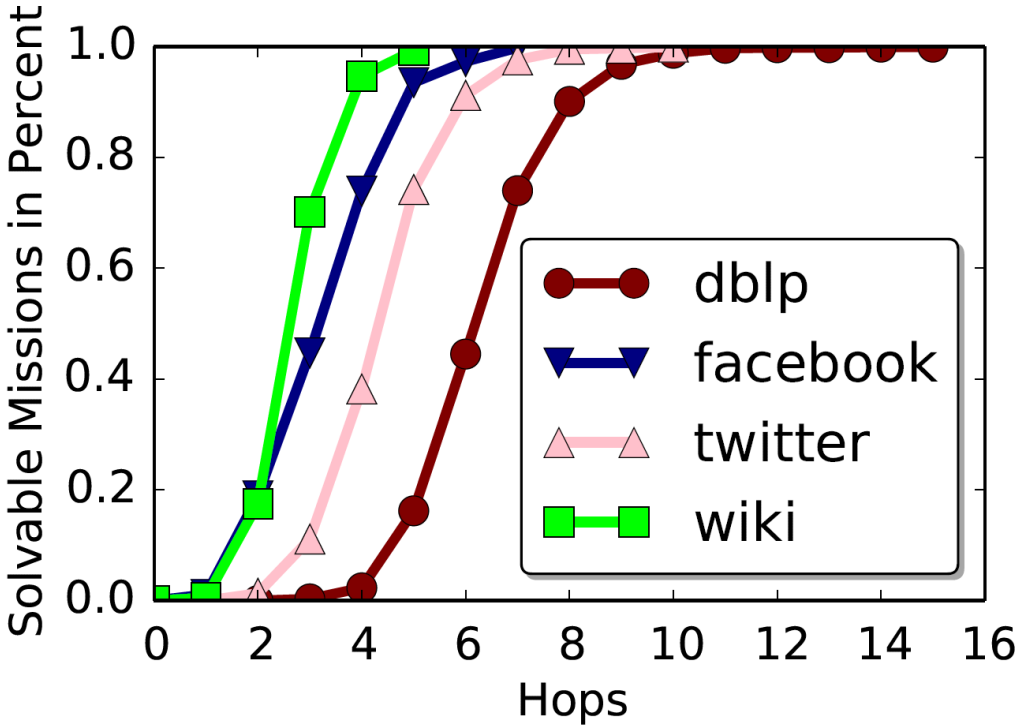
nodes ~4k – ~300k

Experimental Setup & Evaluation

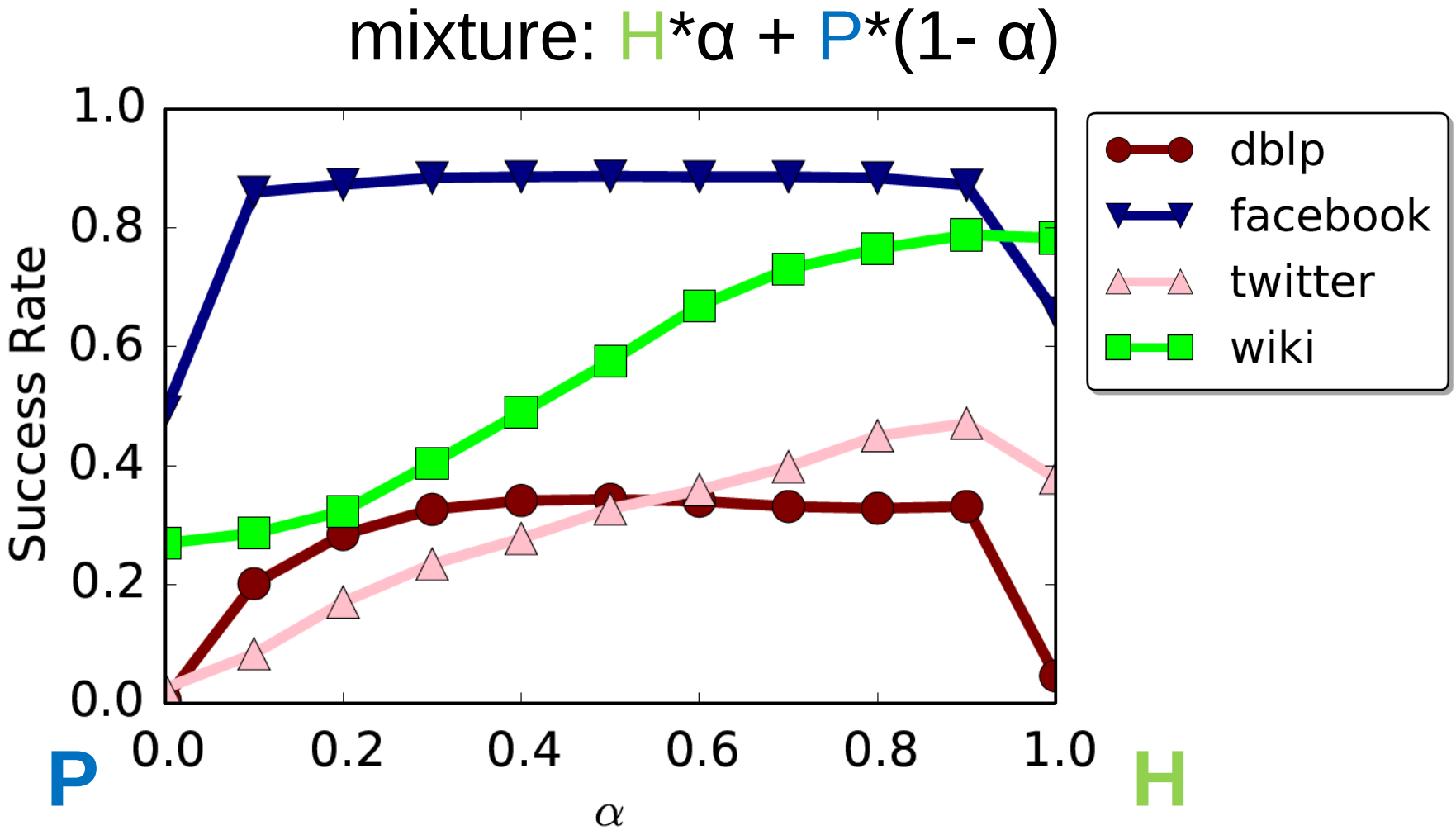
random missions

vary α from 0 to 1

Success Rate

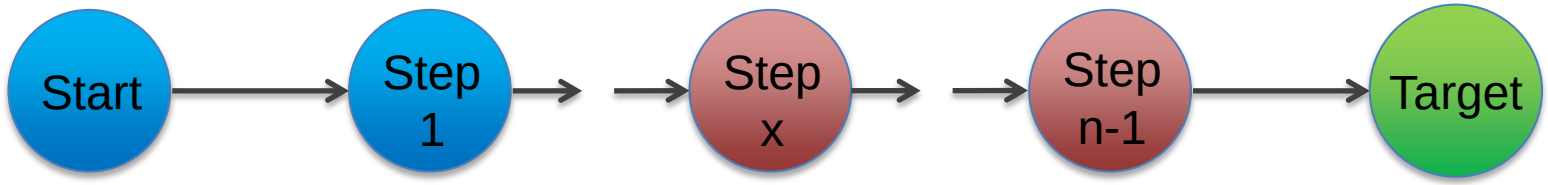


Results Greedy Navigation

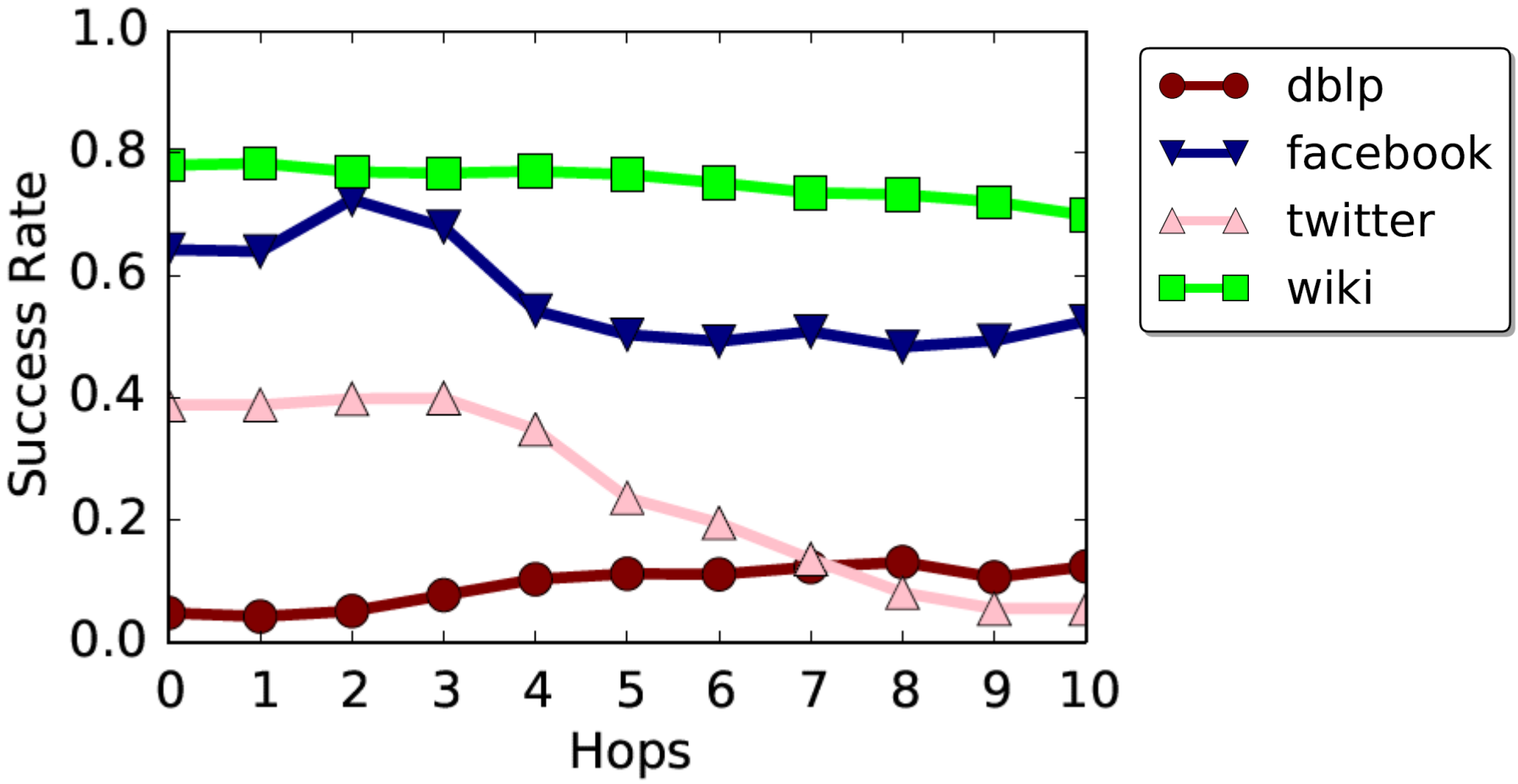


Background Knowledge Models

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

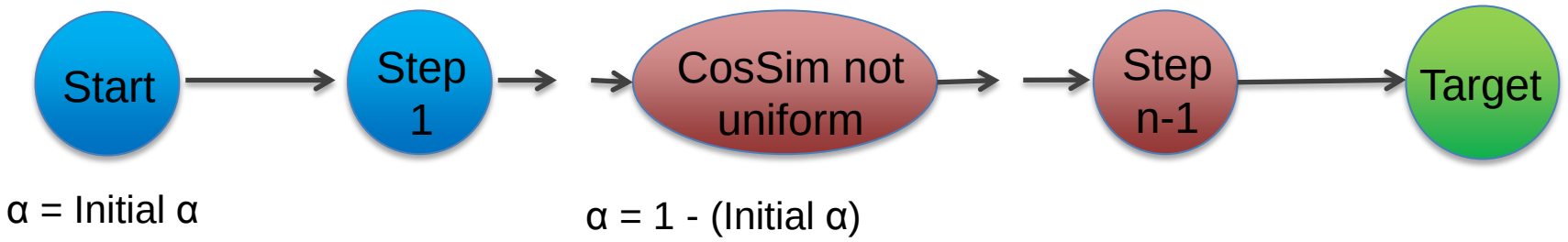


Results Greedy Navigation

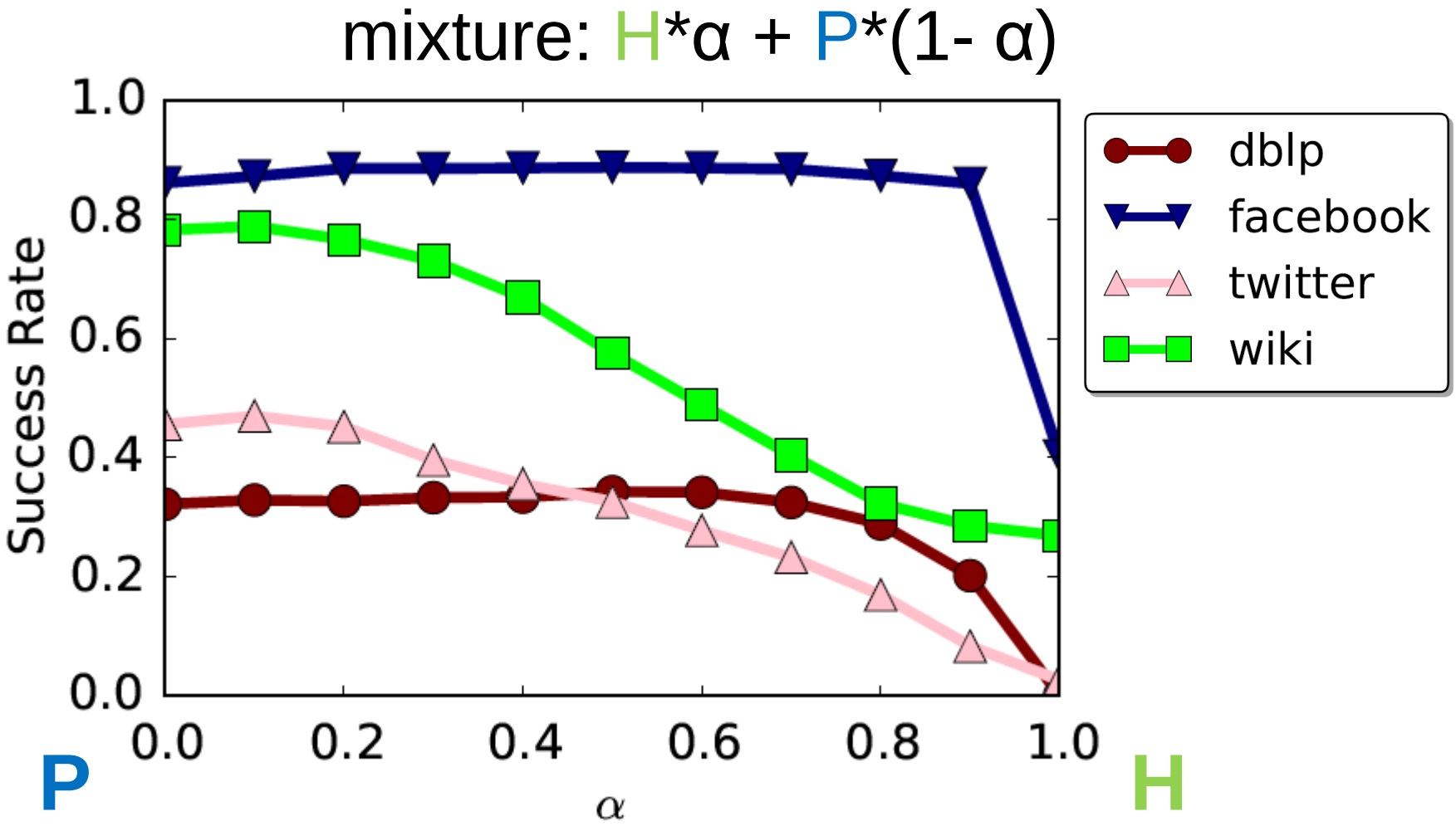


Background Knowledge Models

- dynamic switch



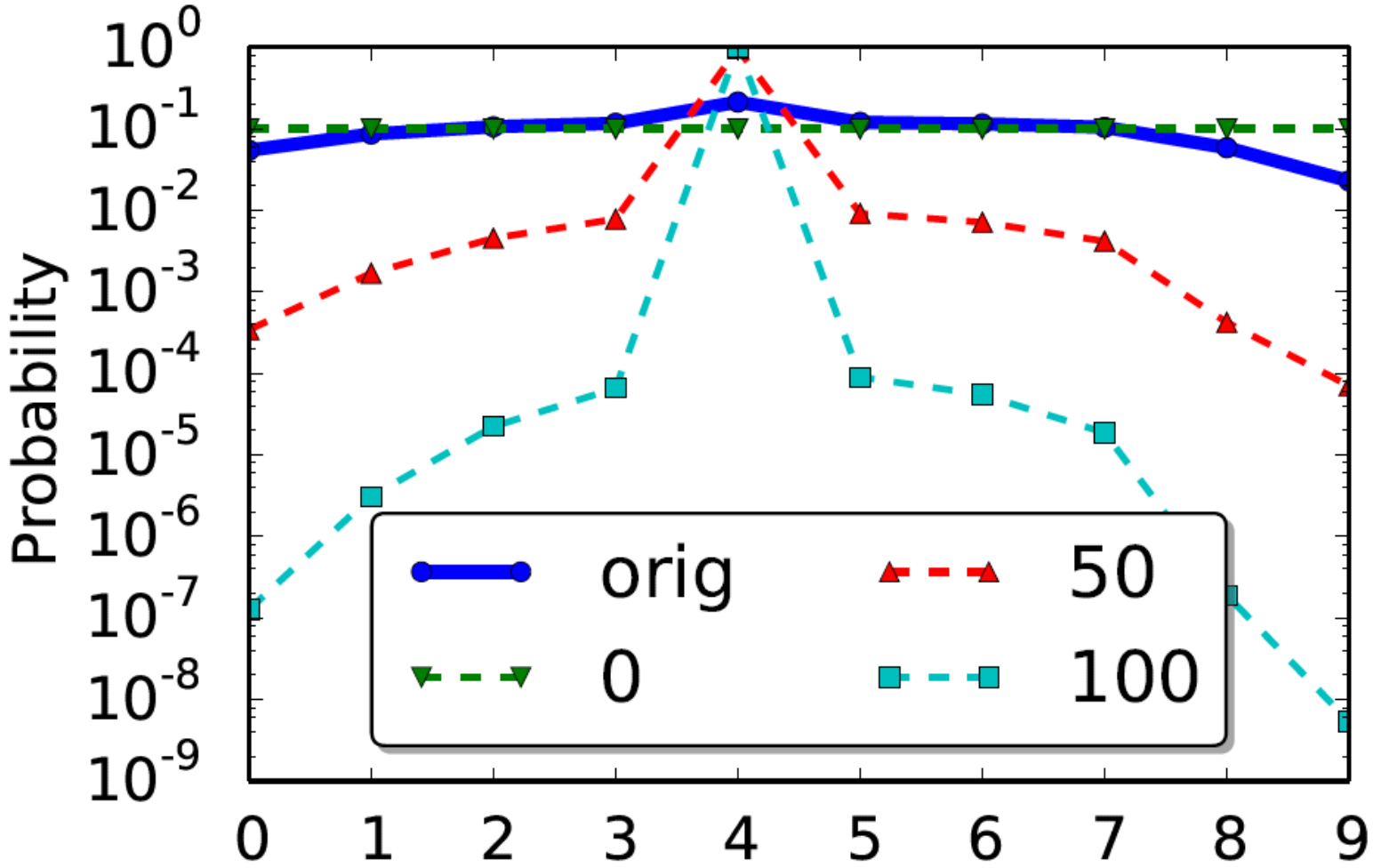
Results Greedy Navigation



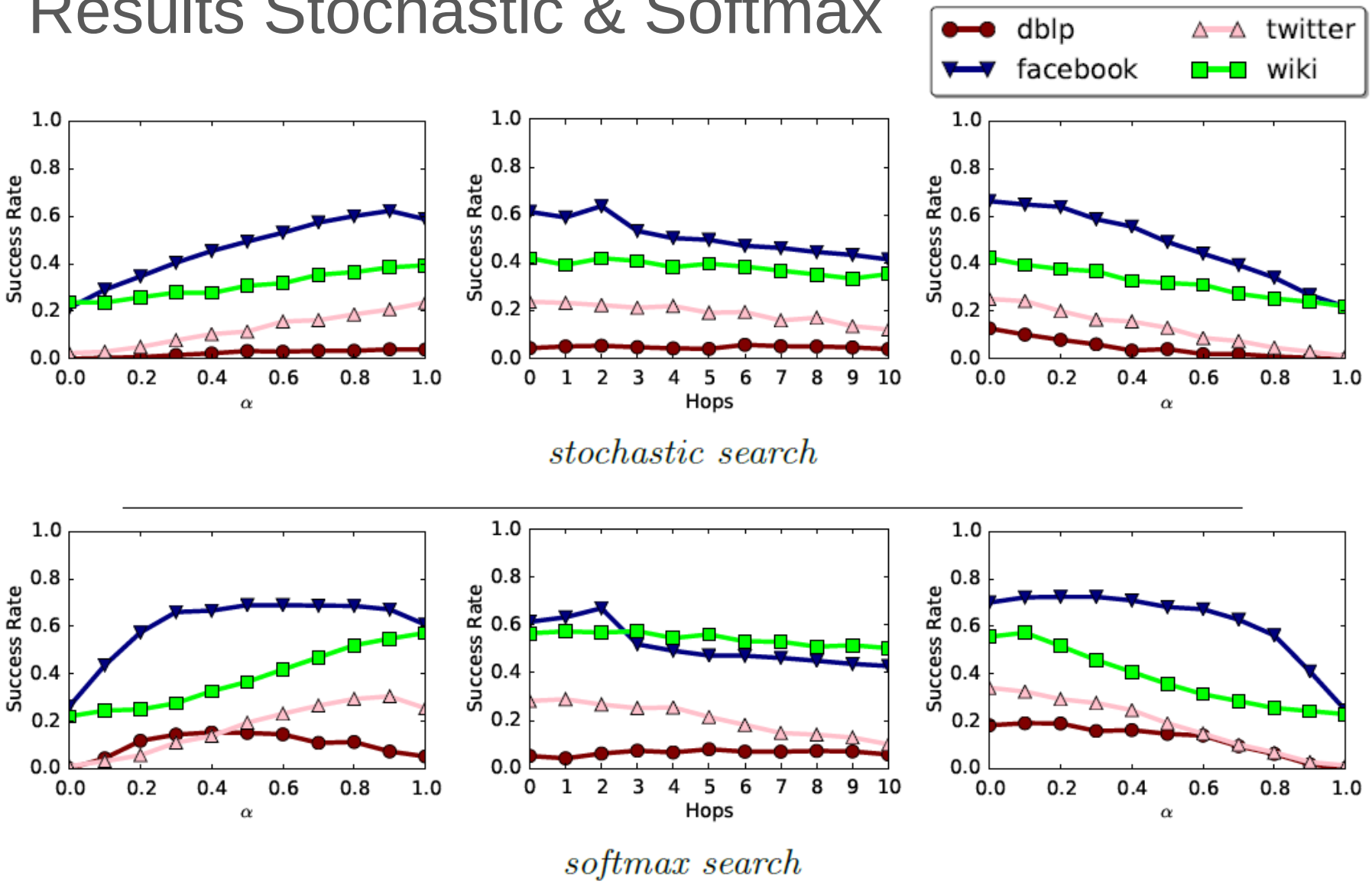
Navigation Models

- 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

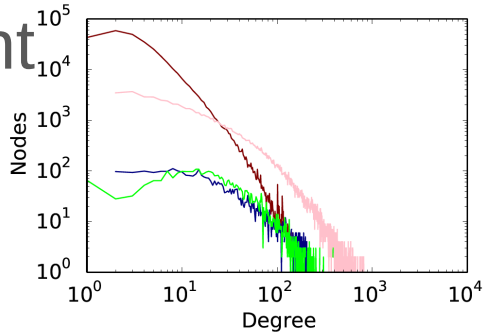

Probability
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
0
1
2
3
4
5
6
7
8
9
orig
50
0
100

Results Stochastic & Softmax

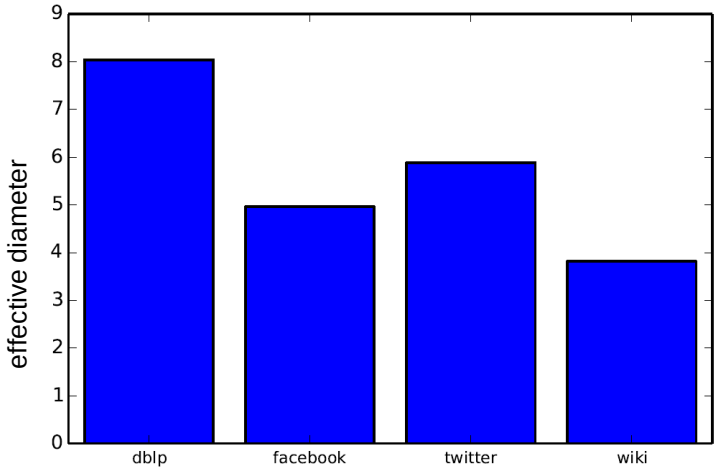


Discussion

- 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 ☺