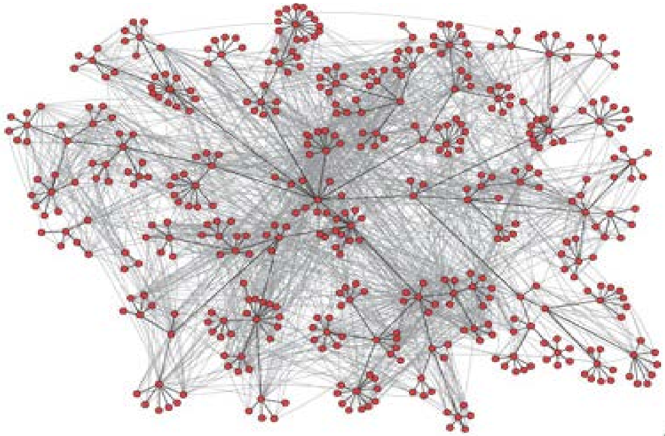
Human Navigation in Networks

Jure Leskovec
Stanford University

Transformation

Online friendships
[Ugander-Karrer-Backstrom-Marlow, '11]



Corporate e-mail communication
[Adamic-Adar, '05]

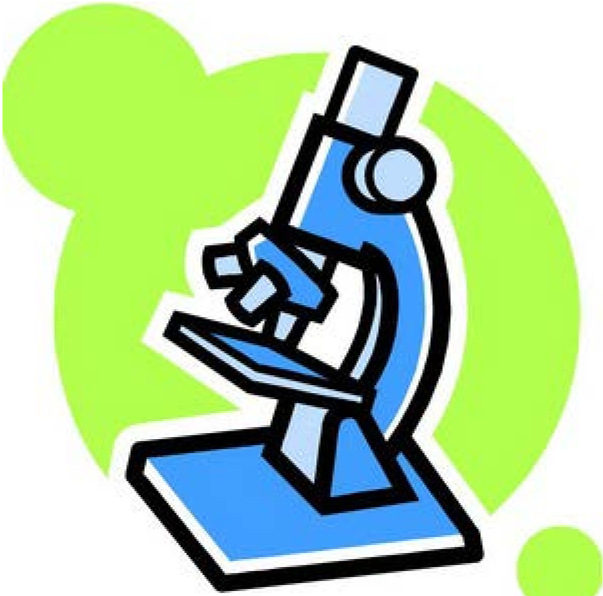
- **Social Transformation of Computing**
  - Technological networks intertwined with social
- **Profound transformation in:**
  - How knowledge is produced and shared
  - How people interact and communicate
  - The scope of CS as a discipline

Two Issues: One

- **Two issues for foundations of computing**
  - **(1) How do we design in this space?**
    Combine social models with core ideas from computing
    - **Complex networks:** design, analysis, models
    - **Algorithmic game theory**: designing with incentives
    - **Social media:** reputation, recommendation, contagion

Two Issues: Two

- **Two issues for foundations of computing**
  - **(2) Science advanced when the invisible becomes visible.**
    - Can we recognize fundamental patterns of human behavior from raw digital traces?
    - Can new computational models address long-standing social-science questions?



Invisible Becomes Visible

- **We are surrounded by linked objects**
  - **Social networks:**
    - Friendships/informal contacts among people
    - Collaboration in companies, organizations, …
  - **Information networks:**
    - Content creation, markets
    - People seeking information
- Traditionally networks were hard to obtain

Invisible Becomes Visible

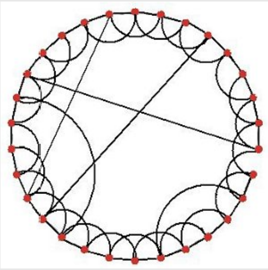
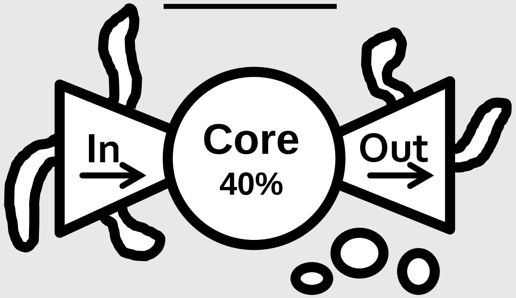
Now: Large on-line systems

- **Social networks:**
  - On-line communities: Facebook, Twitter, …
  - E-mail, blogging, electronic markets
- **Information networks:**
  - Hypertext, Wikipedia, Web

What have we learned about these networks?

Social & Information Networks

- **We know a lot about the structure**

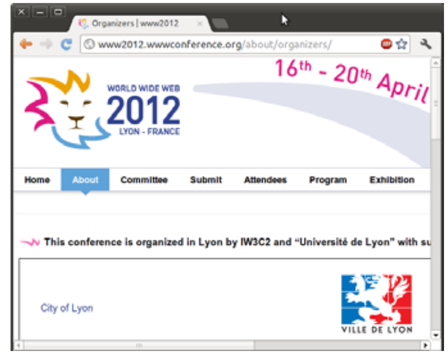
Network Property	Social Networks (MSN [Leskovec,Horvitz '08])	Information Networks (Web [Broder et al. '00])
Connectivity: Well connected	Giant component of 99.9% nodes	Giant component of 90% nodes
Degrees: Heavy-tailed	Log-normal	Power-law
Diameter: Small	6-degrees of separation	~20
Model	**Small-world** 	**Bow-tie** 

Social & Information Networks

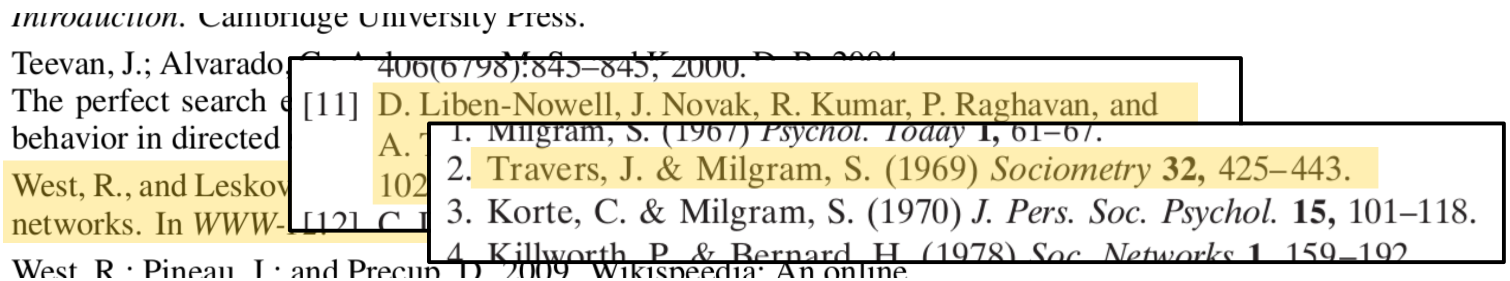
- **We know much less about processes!**
- **What process is common to both?**
- **Navigation!**
  - How people find their way through social networks?
  - How people find information on the Web, Wikipedia?

Navigating Information Nets

- Browsing the Web



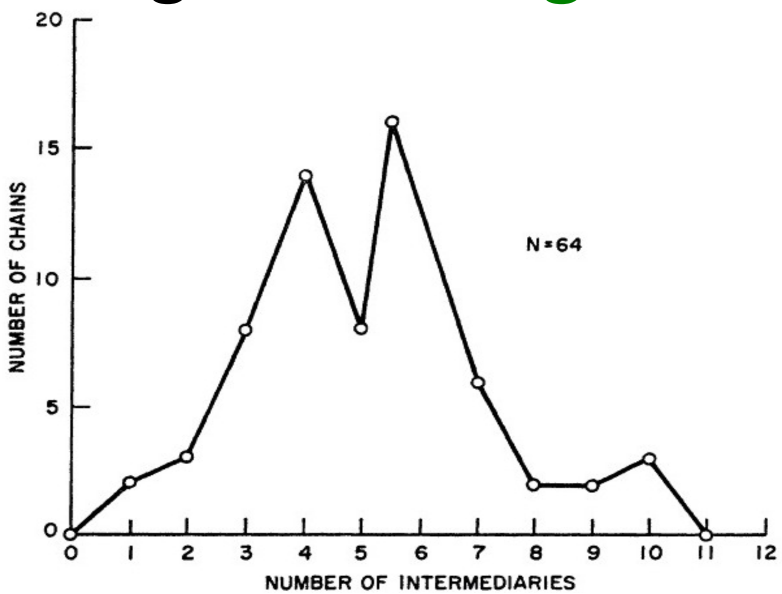
- Literature search



- Consulting an encyclopedia

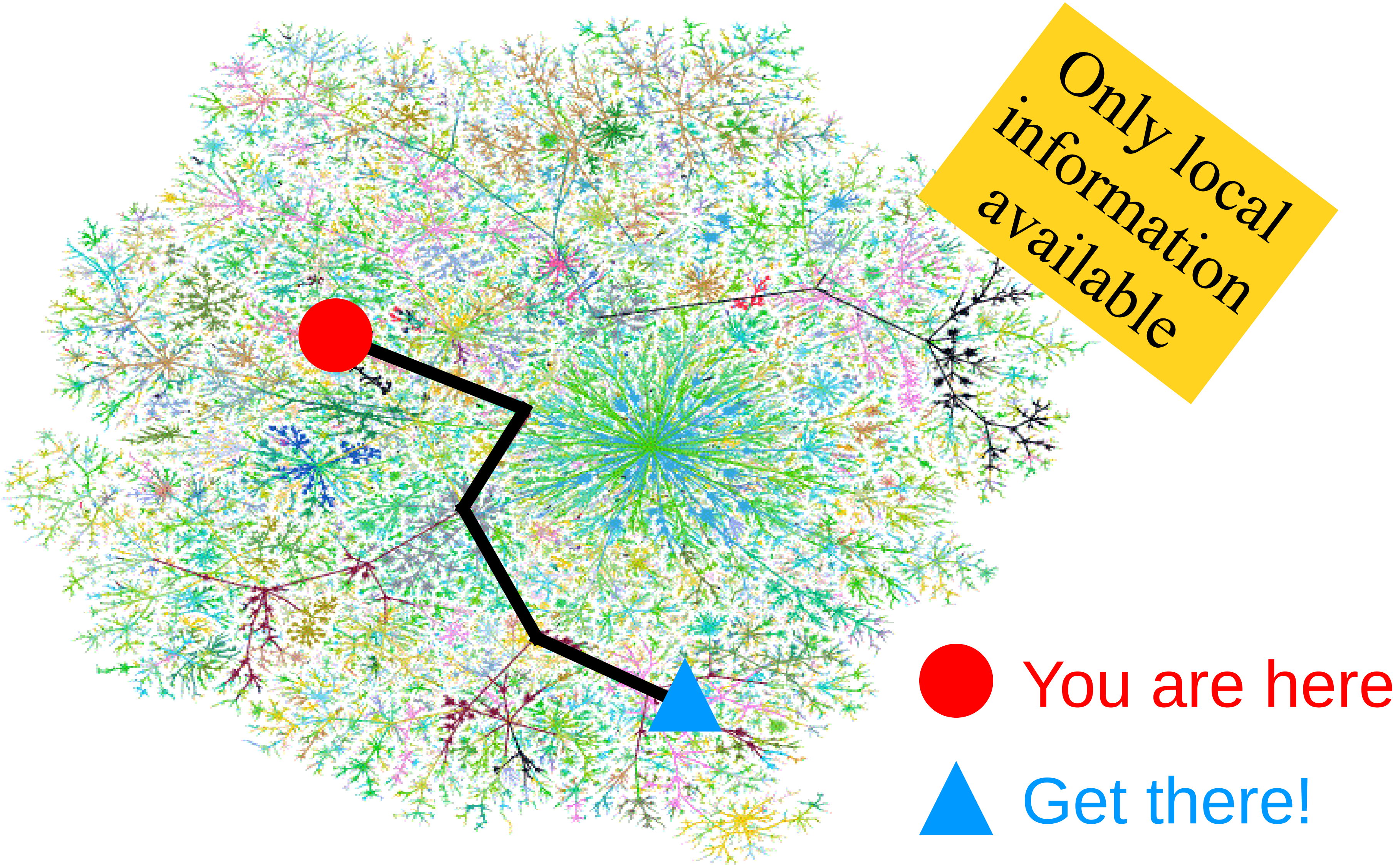
Navigating Social Networks

- **Milgram's small-world experiment** ['67]
  - People forward letters via friends to far-away targets they don't know
  - Six steps on avg. → **Six degrees of separation**



Milgram experiment (Travers-Milgram '70)

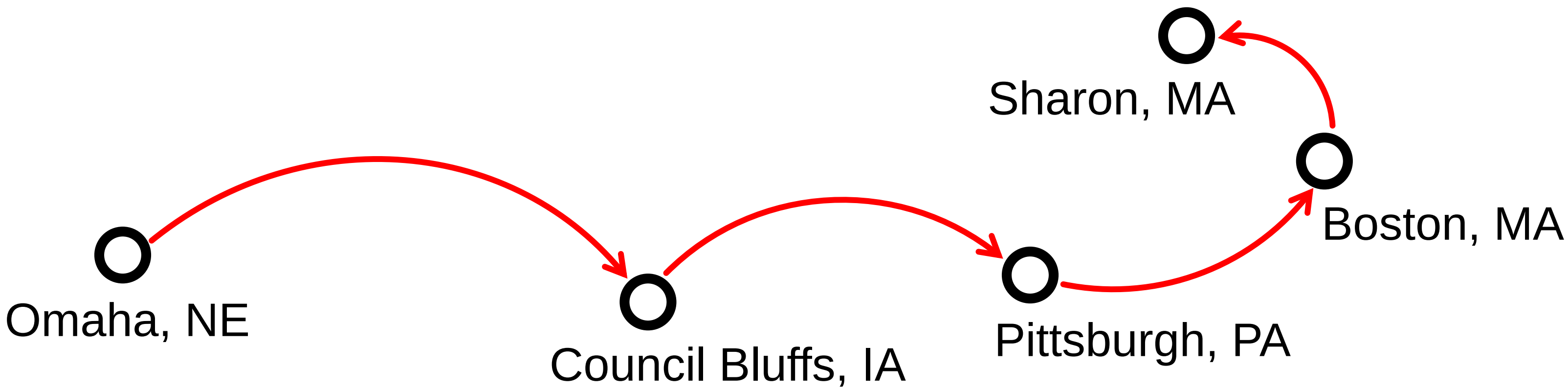
Navigation: Abstraction



Plan for the Talk

- Study **navigation** in **social** as well as **information** networks
  - What is common? What differs?
  - What are the design implications for computing applications and systems?
- **Common theme:**
  Use large-scale online data to as a 'telescope' into these processes

Navigating Social Networks

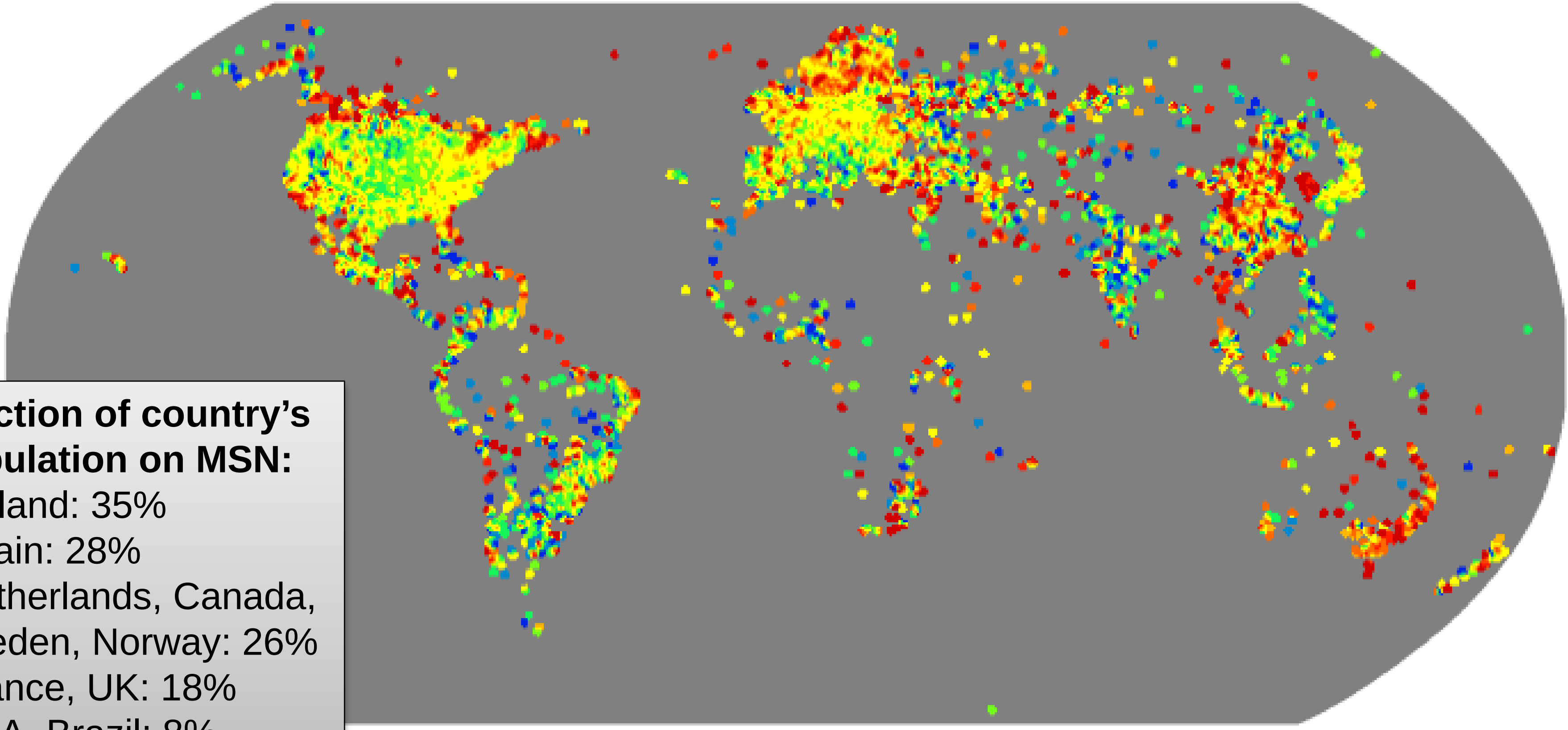


- **Why should strangers be able to find short chains of acquaintances linking them together?**
  - Models for decentralized routing in social networks [Kleinberg '00, Watts-Dodds-Newman '02, ...]

The MSN Network

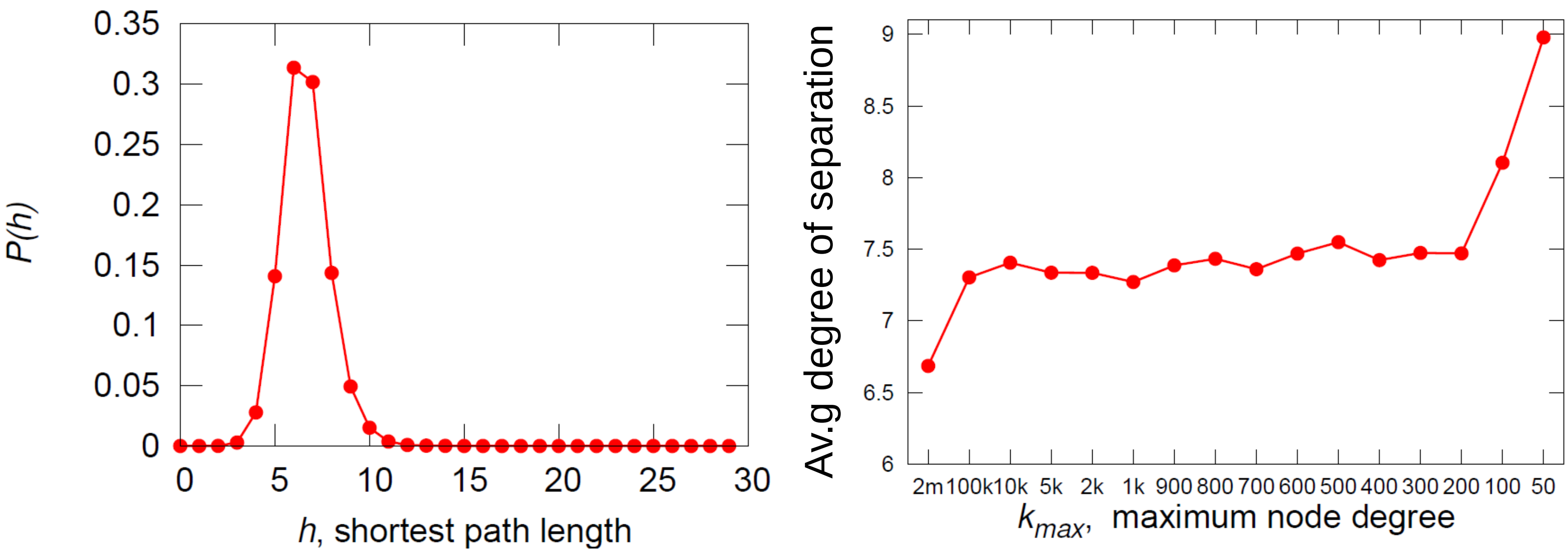
[Leskovec-Horvitz, '08]

- **The MSN Messenger network:**
  - 180 million people, 1.3 billion edges



Fraction of country's population on MSN:
- •Iceland: 35%
- •Spain: 28%
- •Netherlands, Canada, Sweden, Norway: 26%
- •France, UK: 18%
- •USA, Brazil: 8%

6-Degrees of Messaging

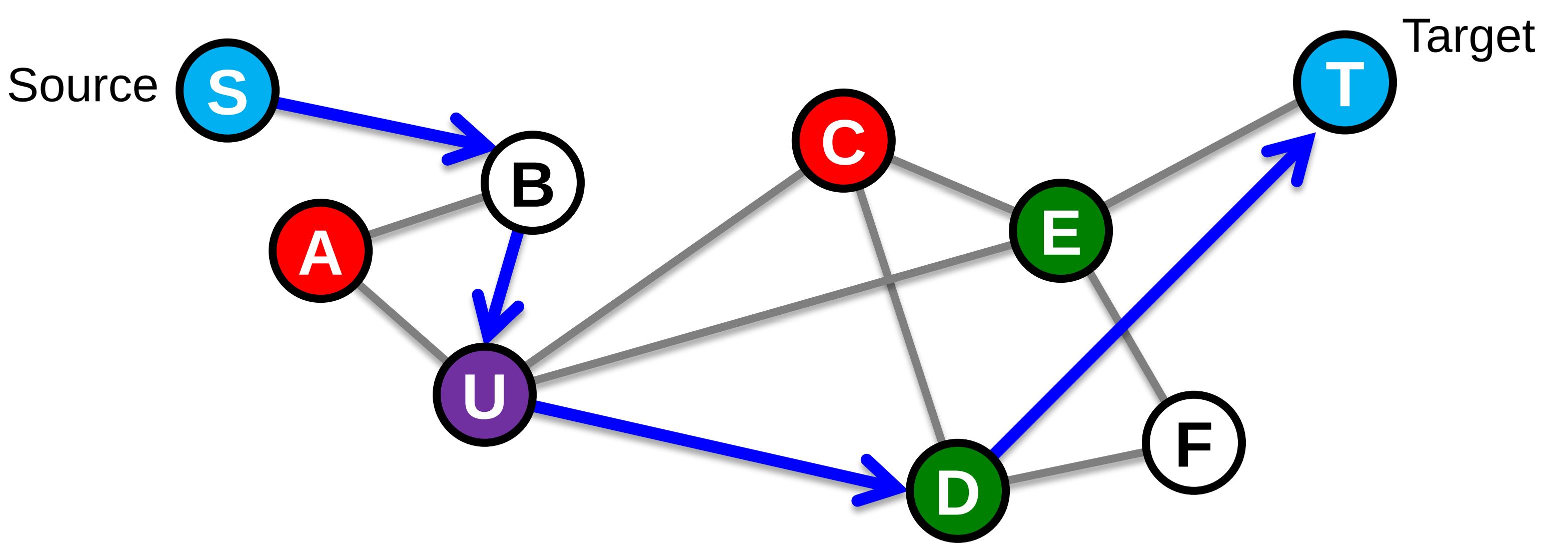


- Avg. degree of separation = 6.6, mode=6
- Long paths (>30) exist in the network
- Network is robust to removal of hubs

[Leskovec-Horvitz, '12]

Navigating the MSN Network

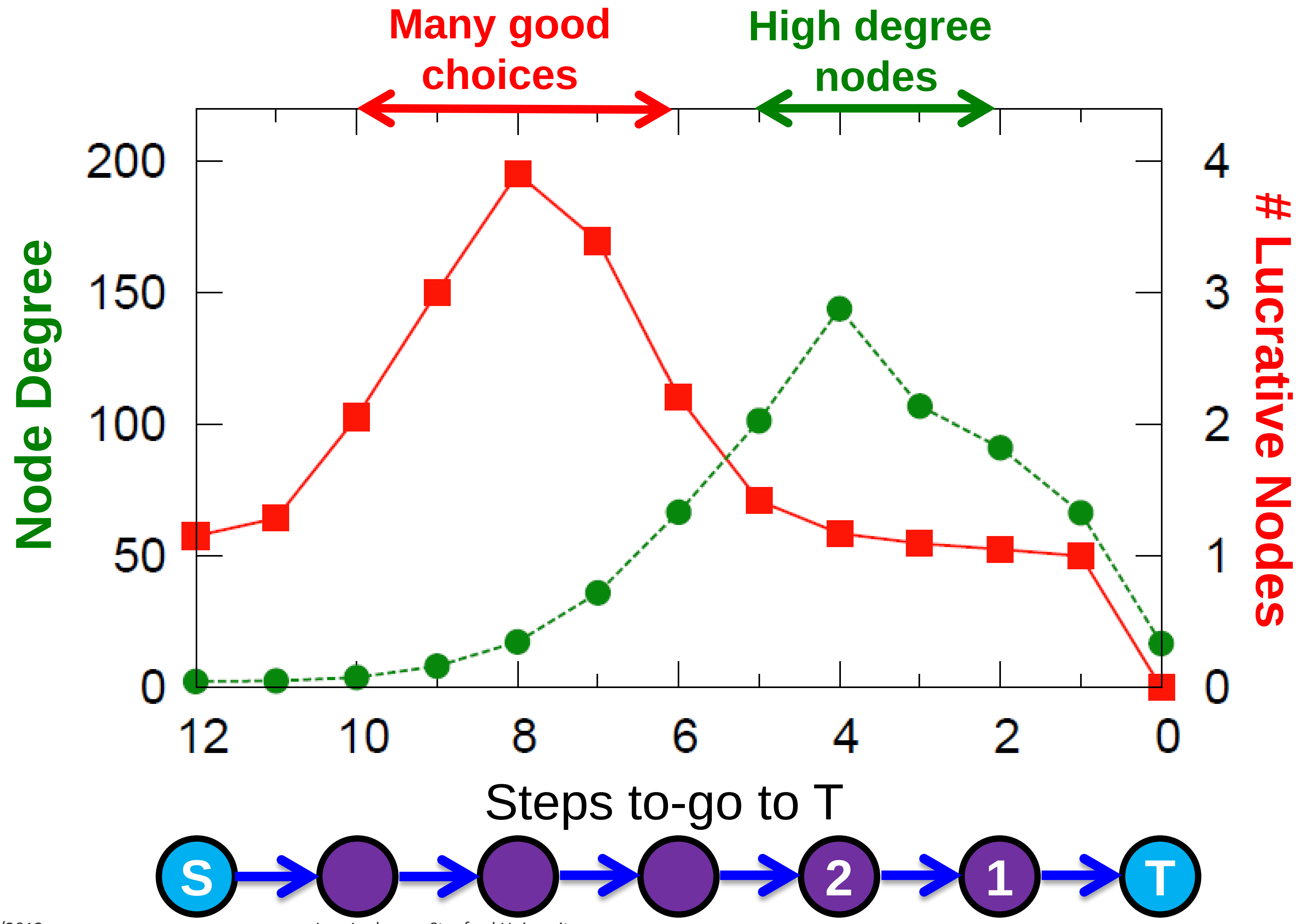
- **What are characteristics of short paths?**
- **How hard is it to find them?**
- **Strategy:** S-T shortest-paths
  - Pick random S-T, run Dijkstra, examine the paths



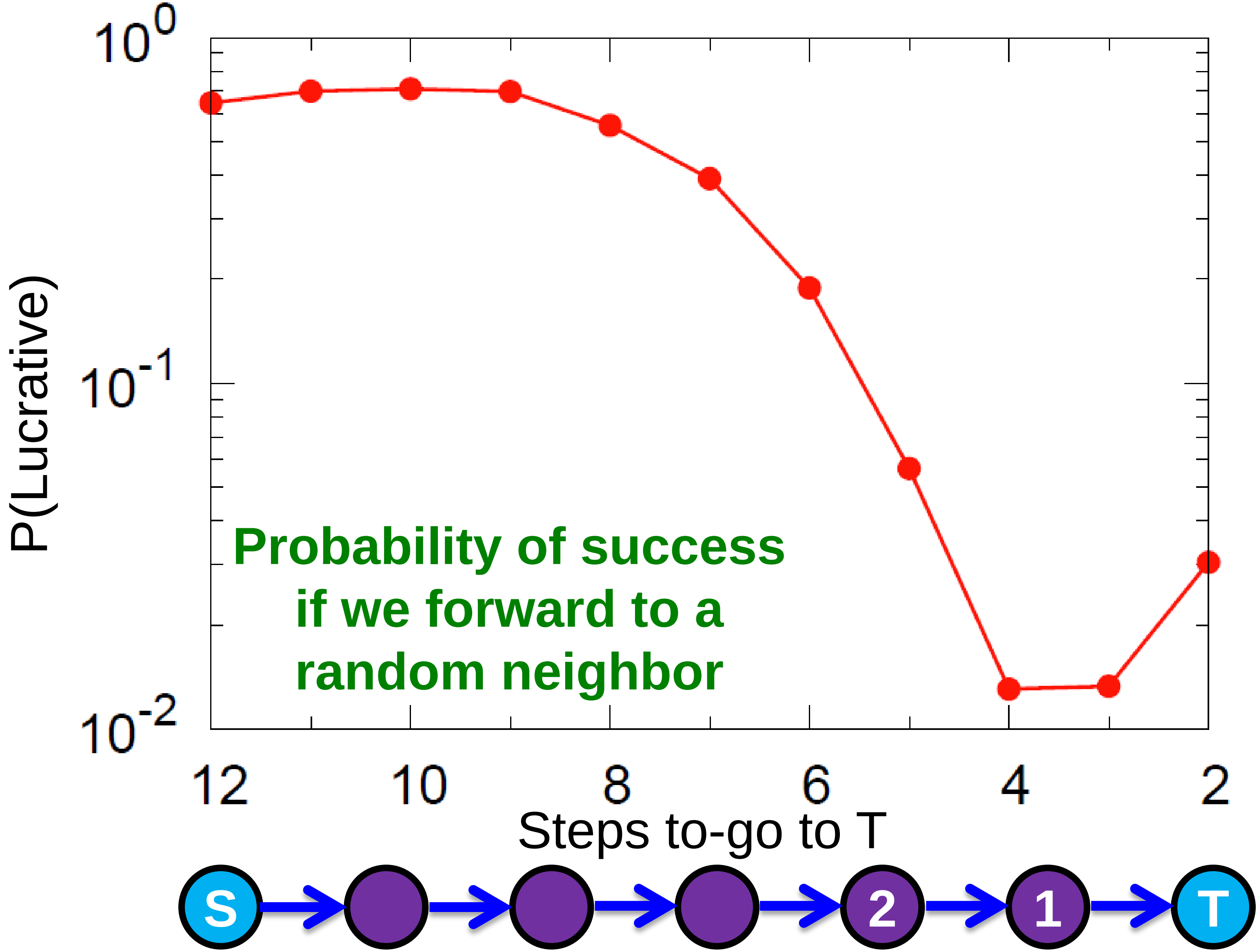
Def: Node is **lucrative**, if it leads "closer" to T

Anatomy of Short Paths

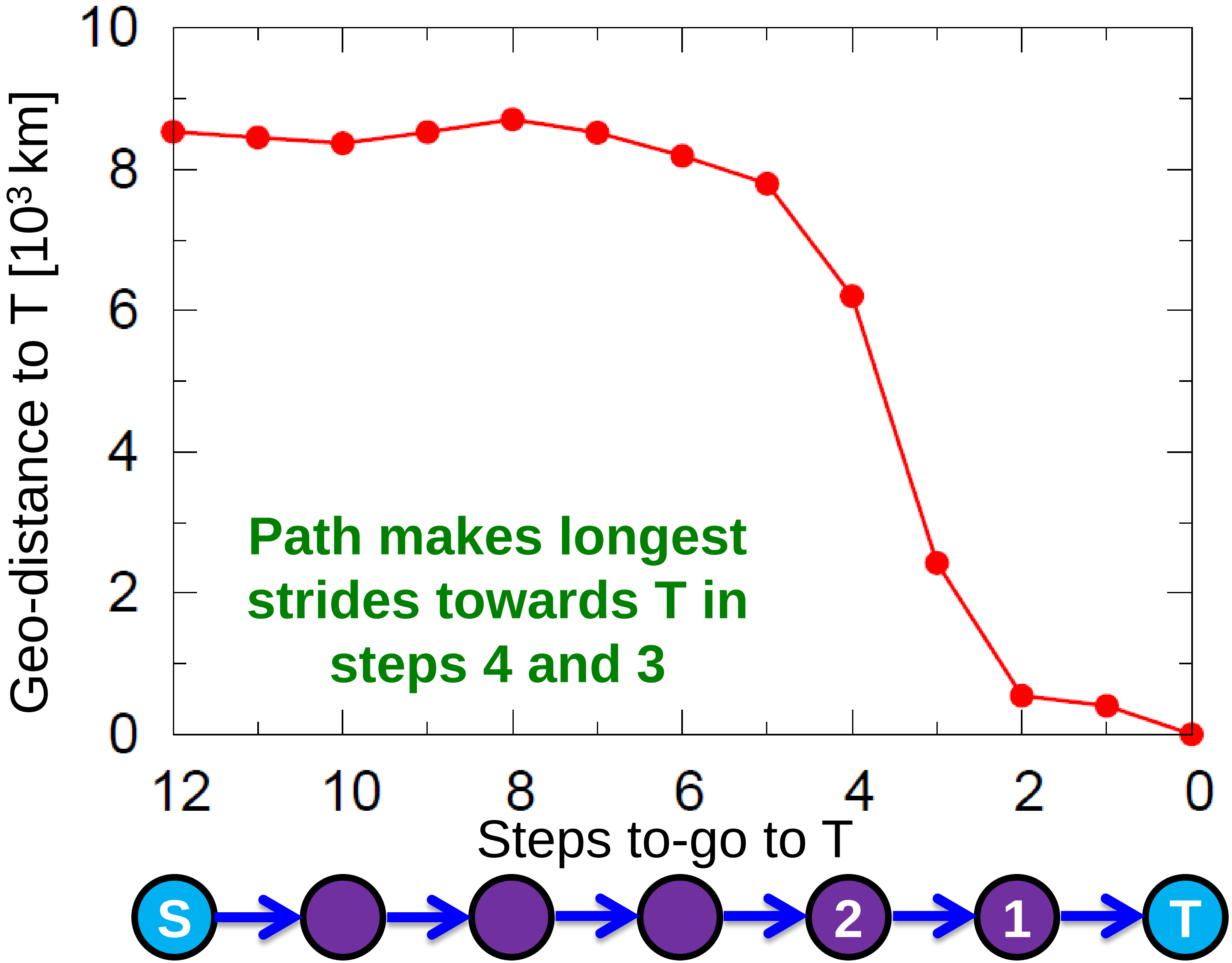
[Leskovec-Horvitz, '12]



Finding a Lucrative Node

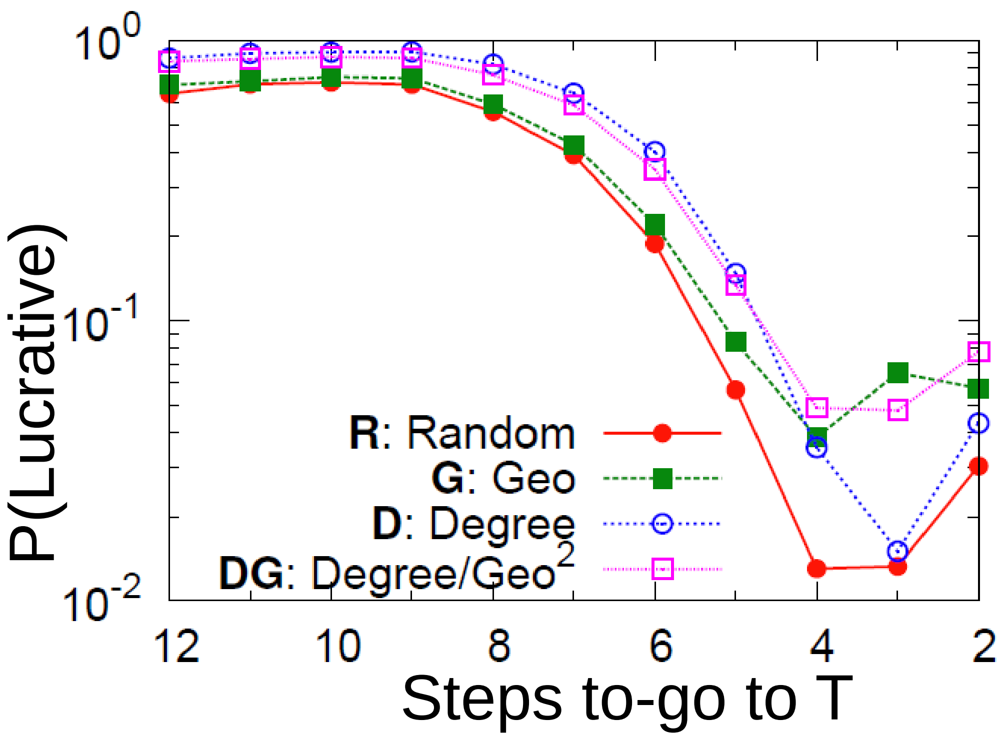


Does geography help?



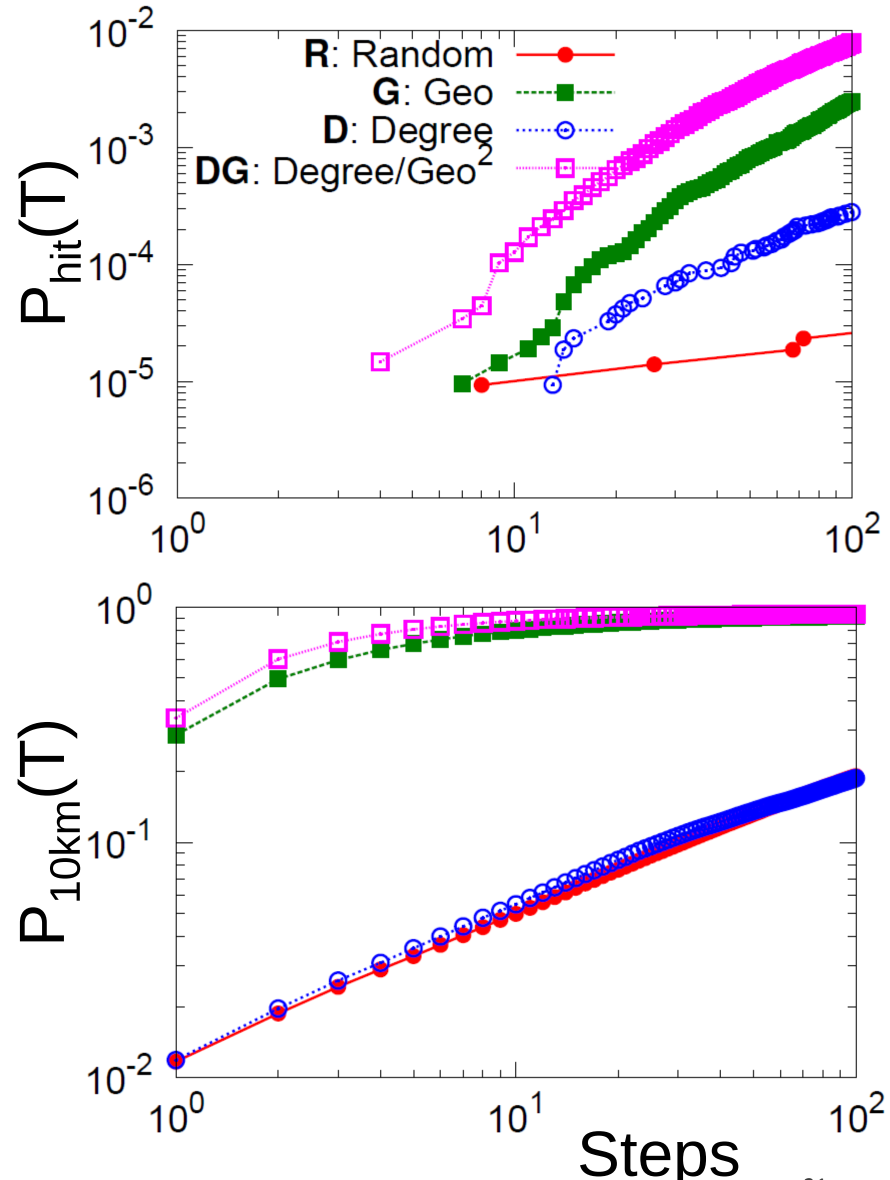
Navigational Heuristics

- **How good are heuristics at navigation?**
- **Heuristics:** Jump to a node X chosen:
  - **R**: Random
  - **G**: $\min \text{geo}(X, T)$
  - **D**: $\max \deg(X)$
  - **DG**: $\min \frac{geo(X,T)}{\deg^2(X)}$



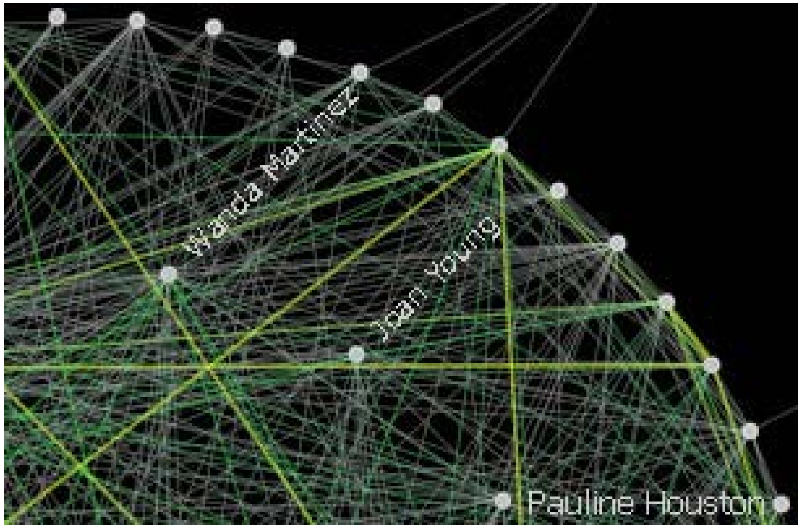
Hitting vs. Getting Close

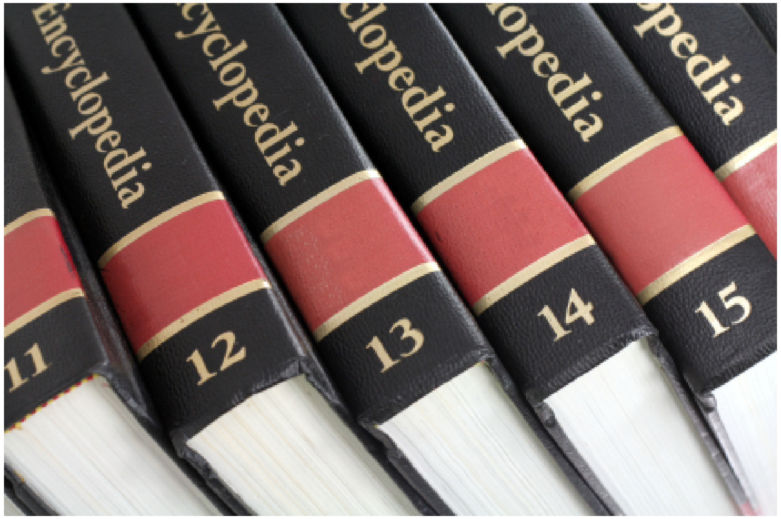
- **Bottom line:**
  - P(hit T in ≤ 10 steps) = 0.001
  - P(get in 10km of T in ≤ 10 steps) = 1
- **Geography provides an important cue but fails in local neighborhoods**



Information Networks

- **How do these translate to navigation in information networks?**
  - Web-browsing
  - Encyclopedia navigation

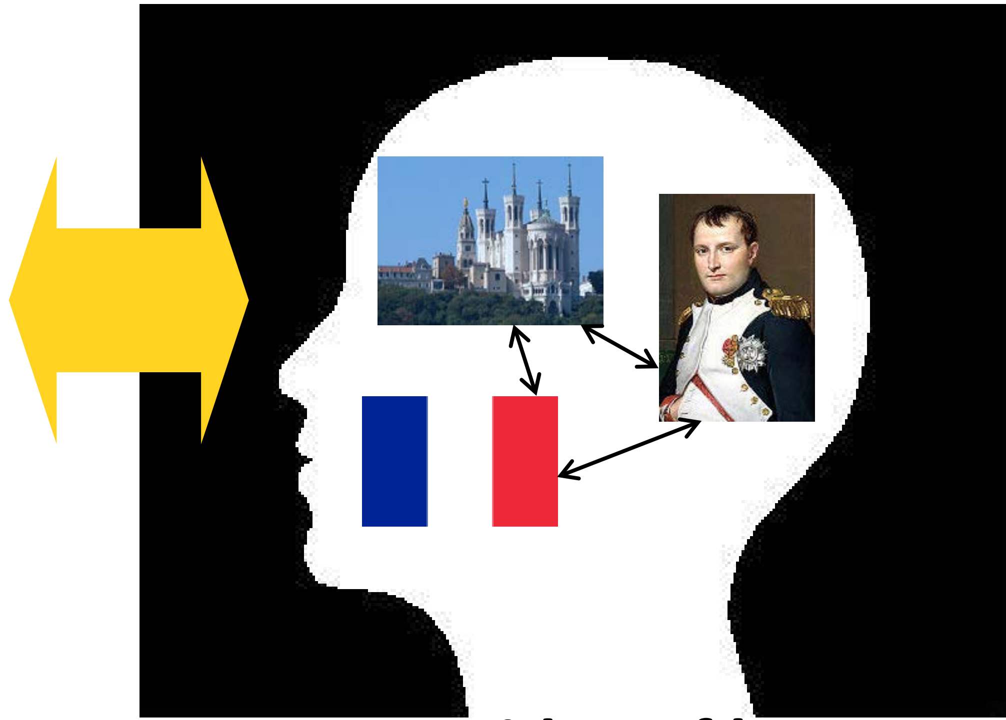




[West-Leskovec, '12]

Navigating Information Nets

- **Large-scale study of navigation in Wikipedia**

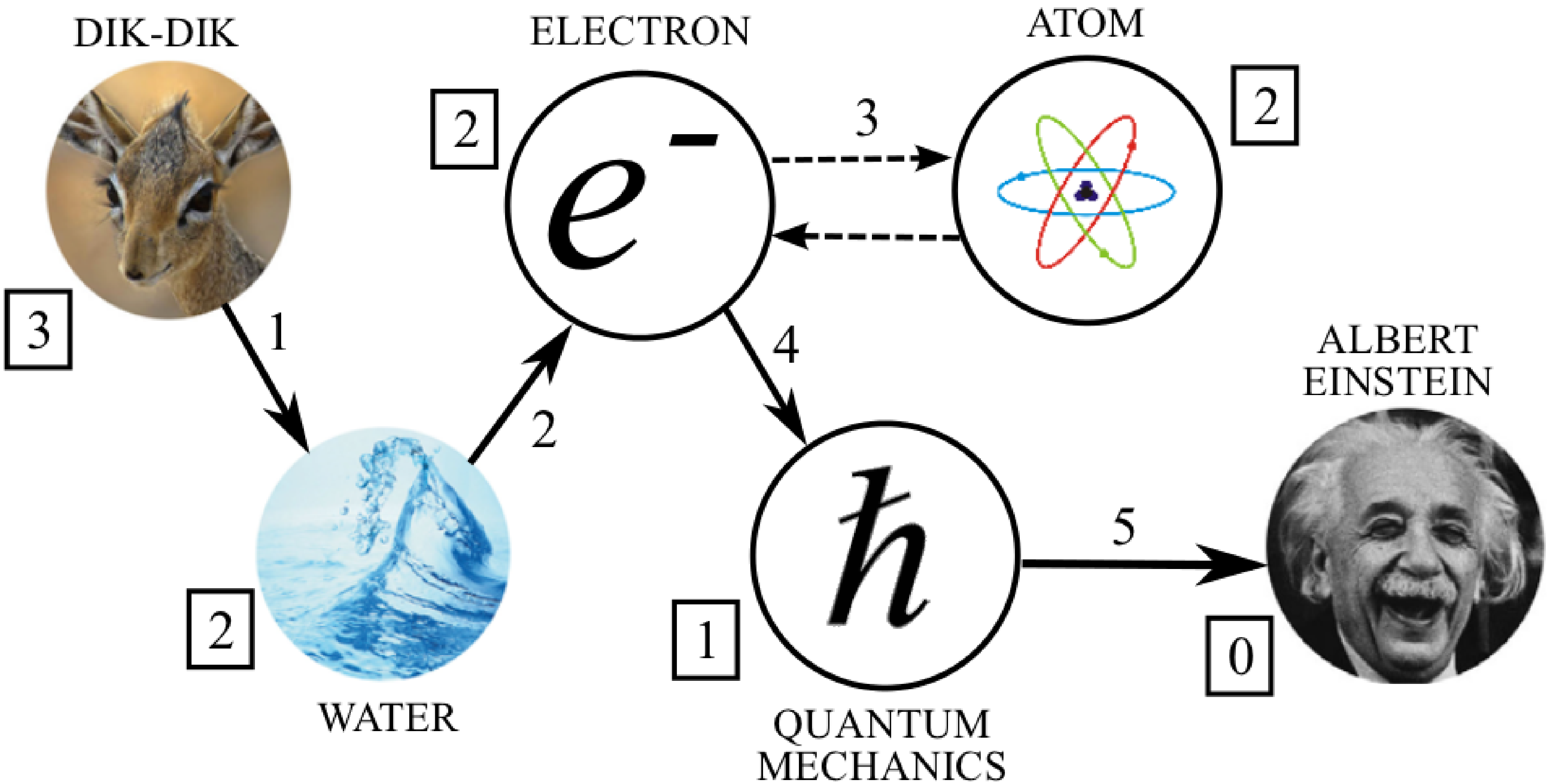


Understand how humans navigate Wikipedia

Get an idea of how people connect concepts

Navigating Wikipedia

[West-Leskovec, '12]



Goal-directed navigation of Wikipedia

Optimal solution: ⟨DIK-DIK, WATER, GERMANY, EINSTEIN⟩

Wikispeedia

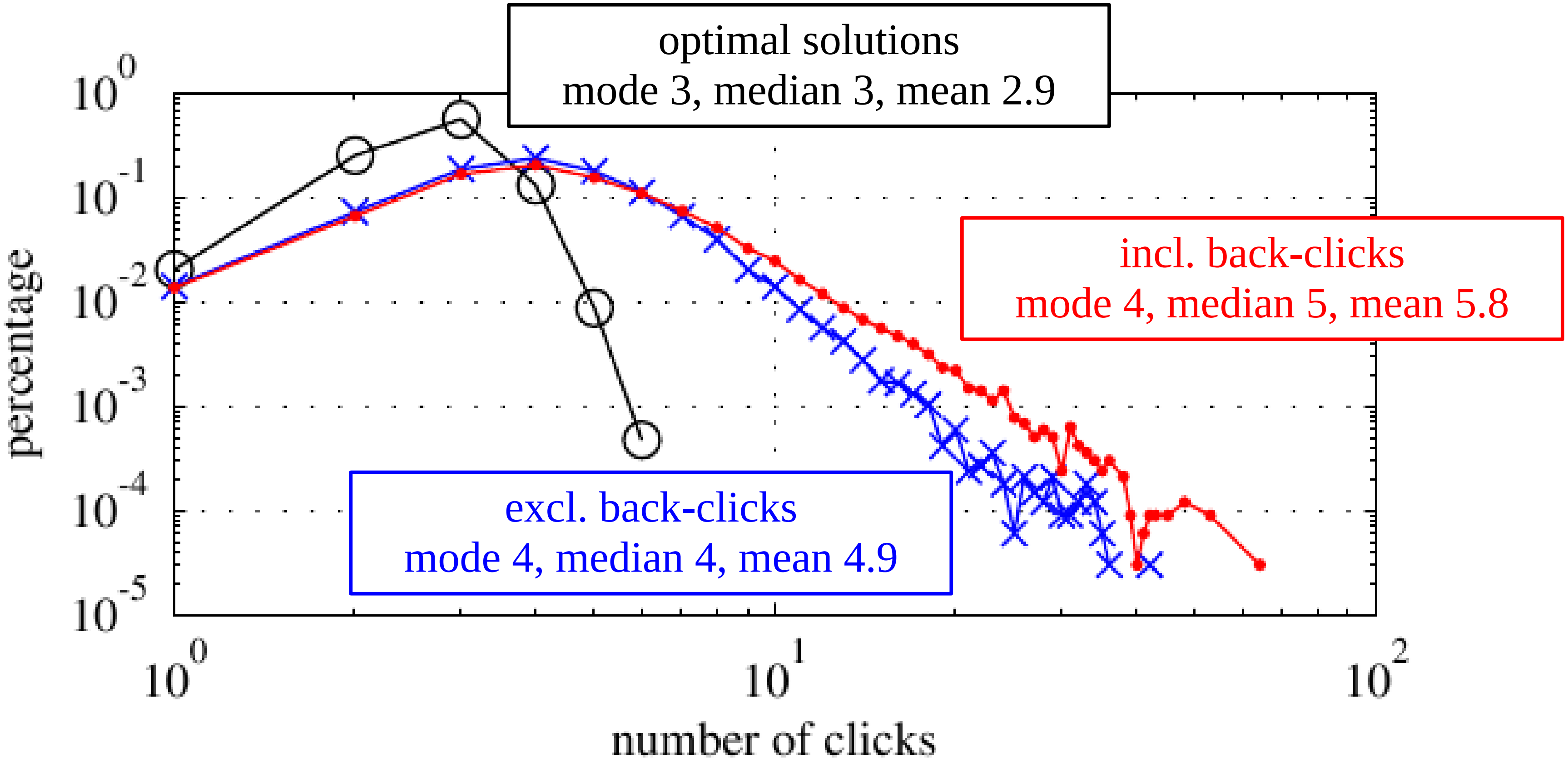
[West-Leskovec, '12]

Wikispeedia: Statistics

- **Graph: "Wikipedia Selection for schools"**
  - 4,000 articles, 120,000 links
  - Shortest paths between all pairs:
    median 3, mean 3.2, max 9
- **Wikispeedia**
  - 30,000 games since Aug 2009
  - 9,400 distinct IP addresses
- **Important:**

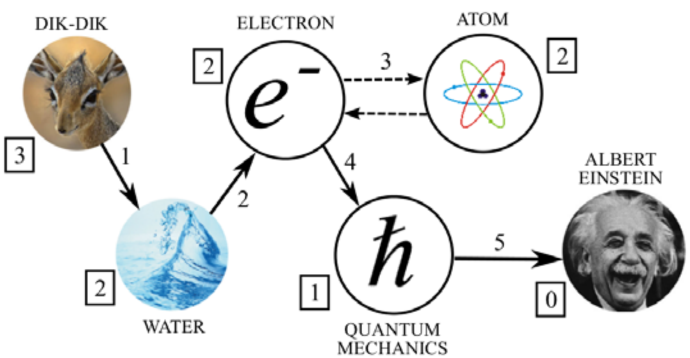
  We know the **target!**

How good are humans?

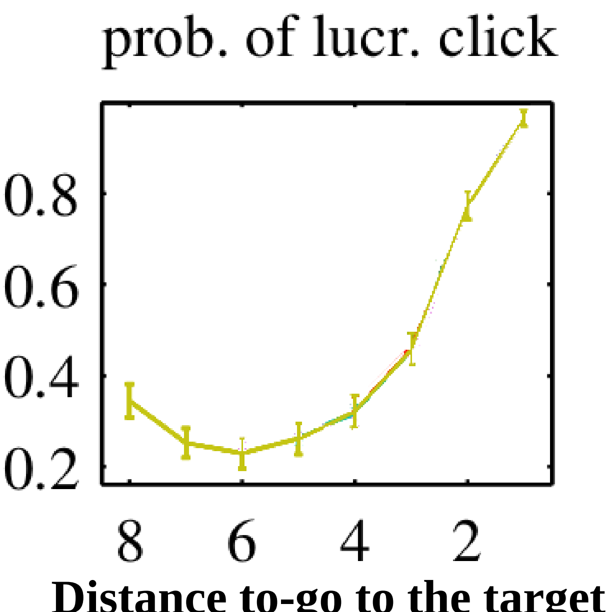


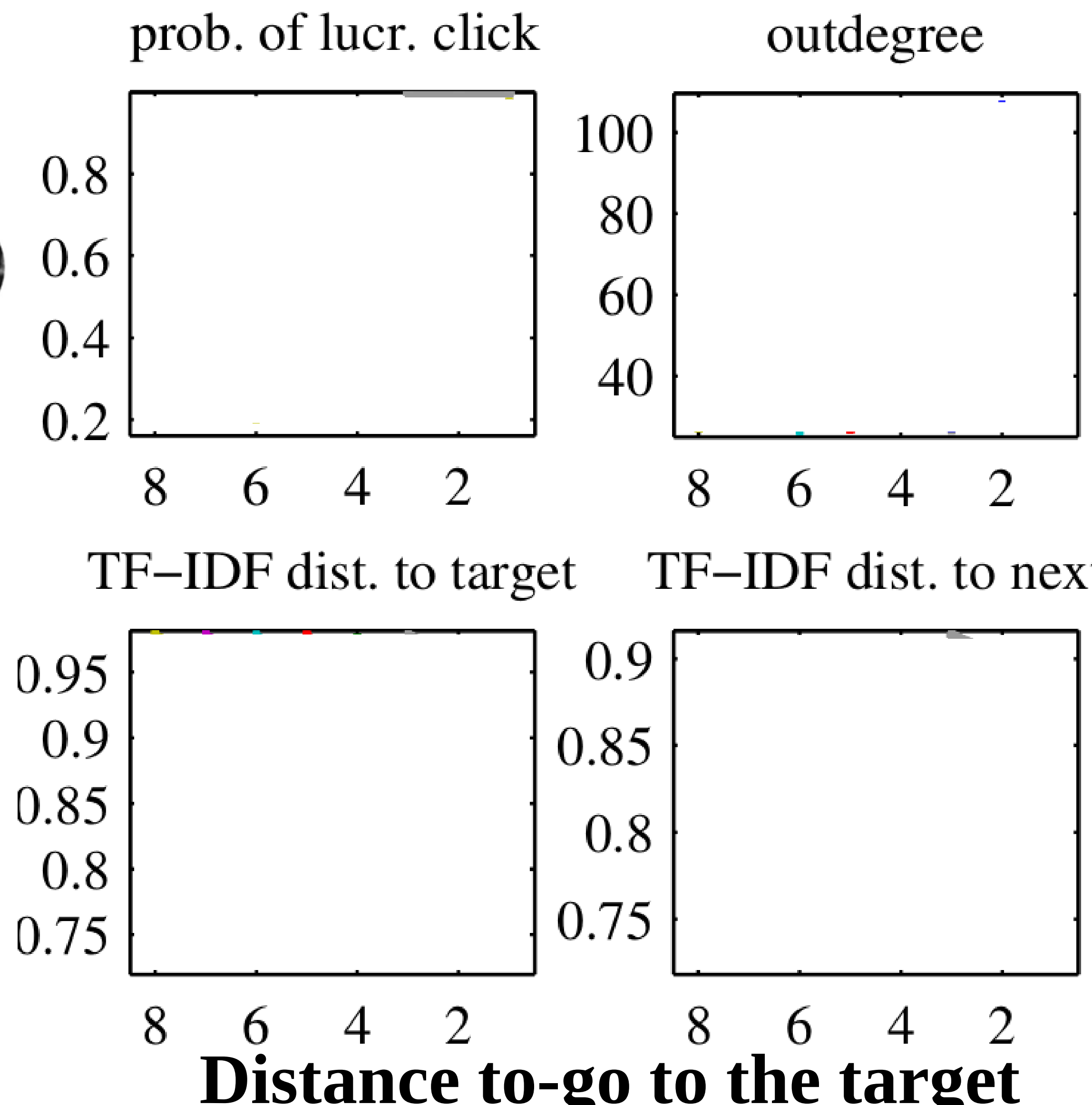
- Larger variance in human than opt. paths
- Overall, humans not much worse than opt.

Anatomy of Wayfinding



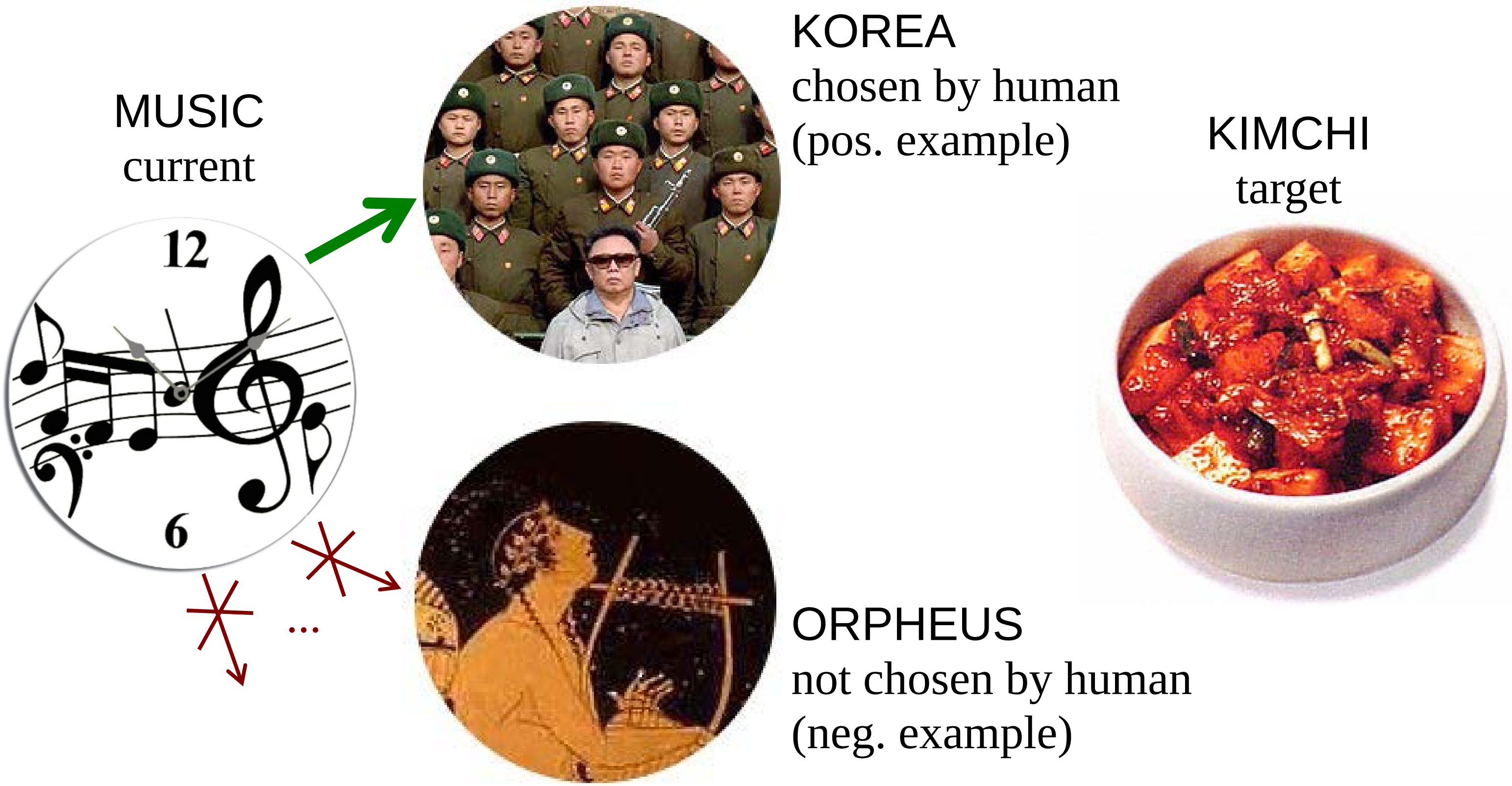
Only missions of SPL 3





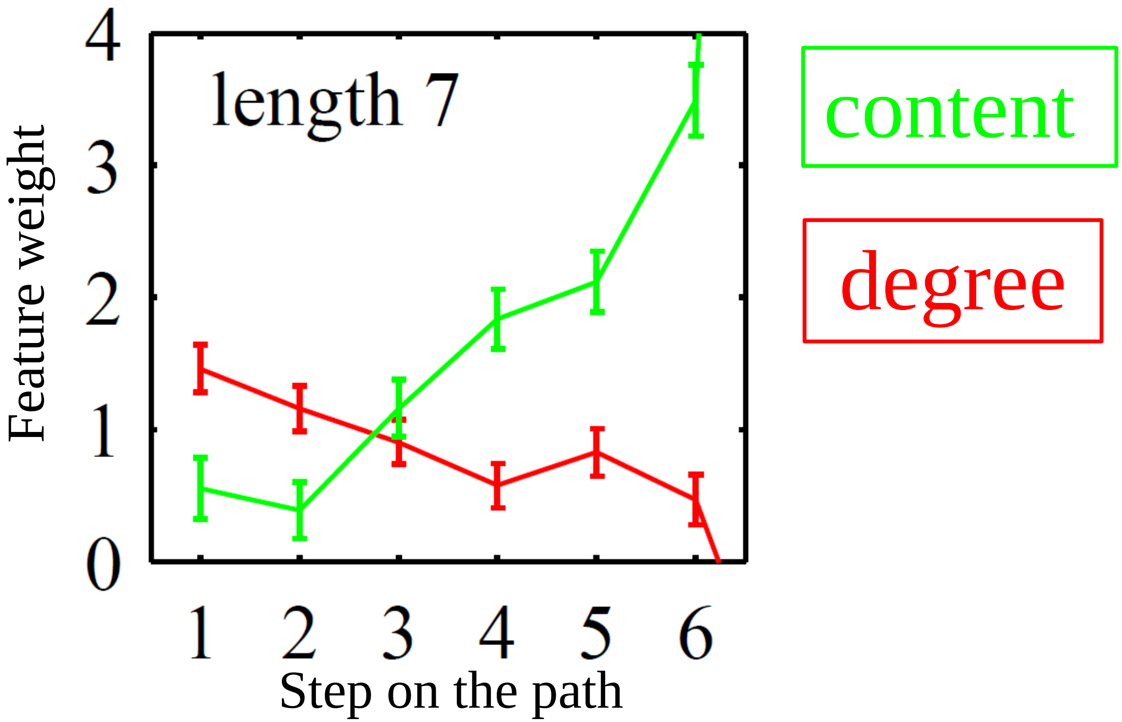
Content vs. the Network

- **For each path position:**
  - Logistic regression to predict human choice



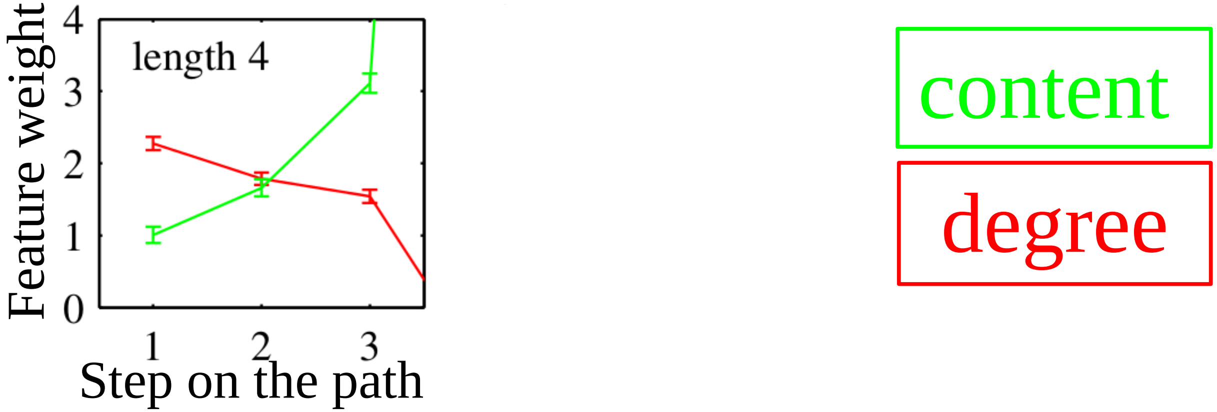
Content vs. the Network

- **For each path position:**
  - Logistic regression to predict human choice
  - Inspect weights for content similarity & degree



Content vs. the Network

- **For each path position:**
  - Logistic regression to predict human choice
  - Inspect weights for content similarity & degree



Endgame Strategies

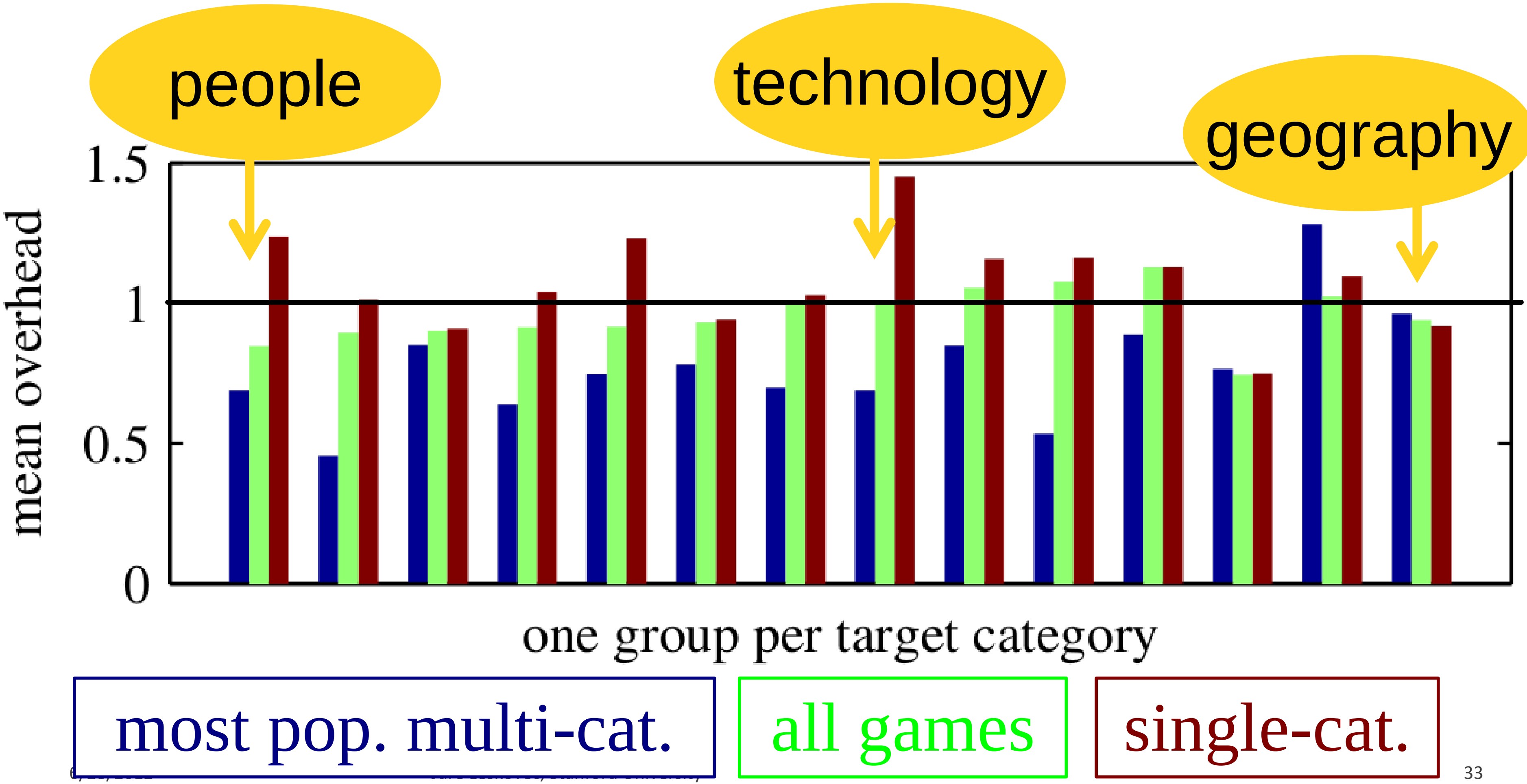
- **Path:**

  … → Water → Germany → Albert Einstein
- **Endgame strategy:**
  - Map last 3 articles to categories:

    Science → Geography → People
- Few popular endgame strategies
  - (Target category)^3 typically most popular
- Among non-target categories, Geography most popular

Endgame Efficiency

$$\textbf{Overhead} = \frac{\text{human game length} - \text{optimal game length}}{\text{optimal game length}}$$



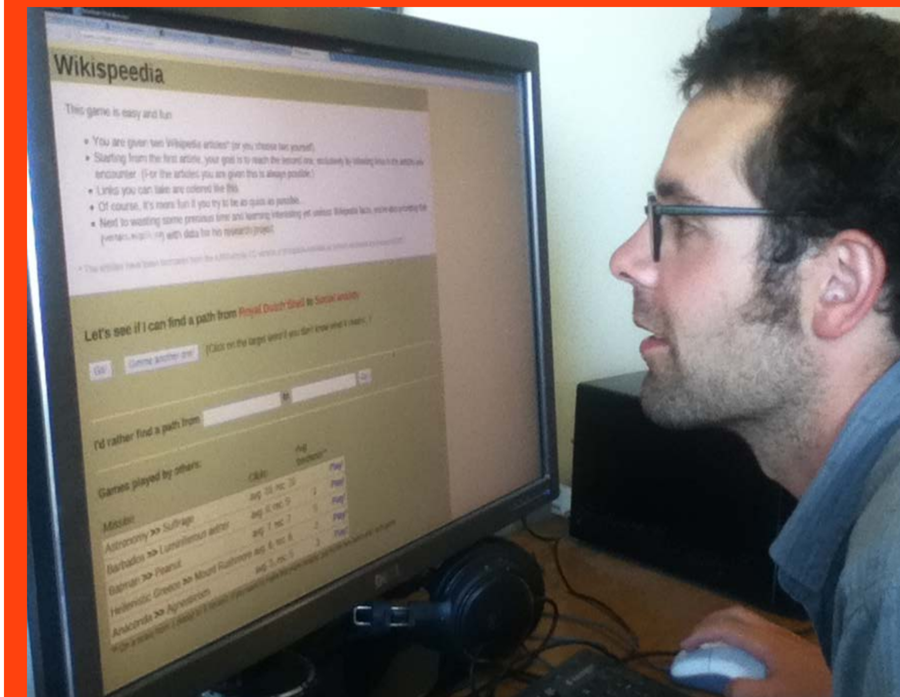
Machine vs. Human

Can we build machines that navigate better than humans?

Machine vs. Human

*WikispeediaAgent.java

```
public Result play(String start, String goal) {
    while (!stack.empty()) {
        // take node from stack:
        s = stack.pop();
        // remember when we've visited the curr
        nodeSequence.add(s);
        nodeSet.add(s);
        // we found what we're looking for, so
        // from the predecessor record:
        if (s.equals(goal)) {
            return new Result(nodeSequence, pre
                    lucrativeClicks, goalDistHi
        }
        // find the successors that haven't bee
        // Set<String> succ = Sets.diff(index.g
        Set<String> succ = Sets.diff(index.getS
        // no unvisited successors; this means
```



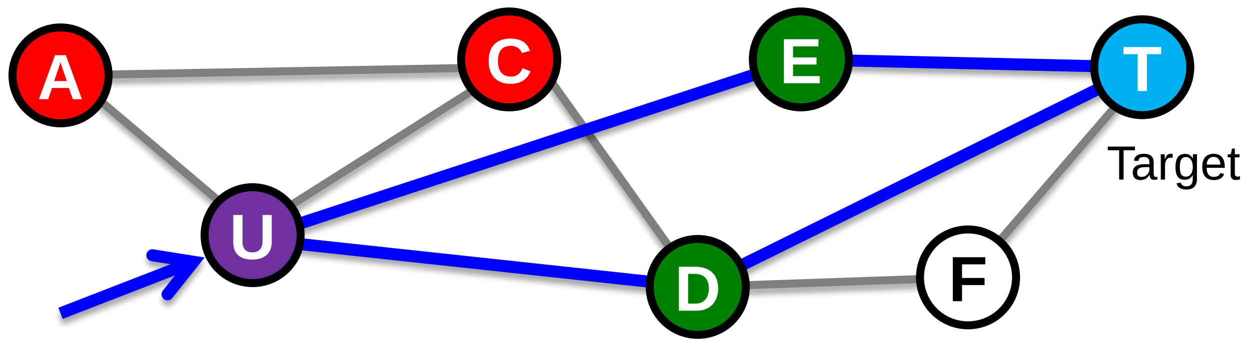
No common sense, only low-level knowledge such as word counts

Common sense and high-level background knowledge

Who is better?

Decentralized Agents

- An agent aims to navigate to target T



- Agent is currently at node U and navigates to neighbor W s.t.

$$W = \arg\max_{U \to W} V(W|U,T)$$

  - Ideally: $V(E|U,T) > V(C|U,T)$

- **What is the value function?**

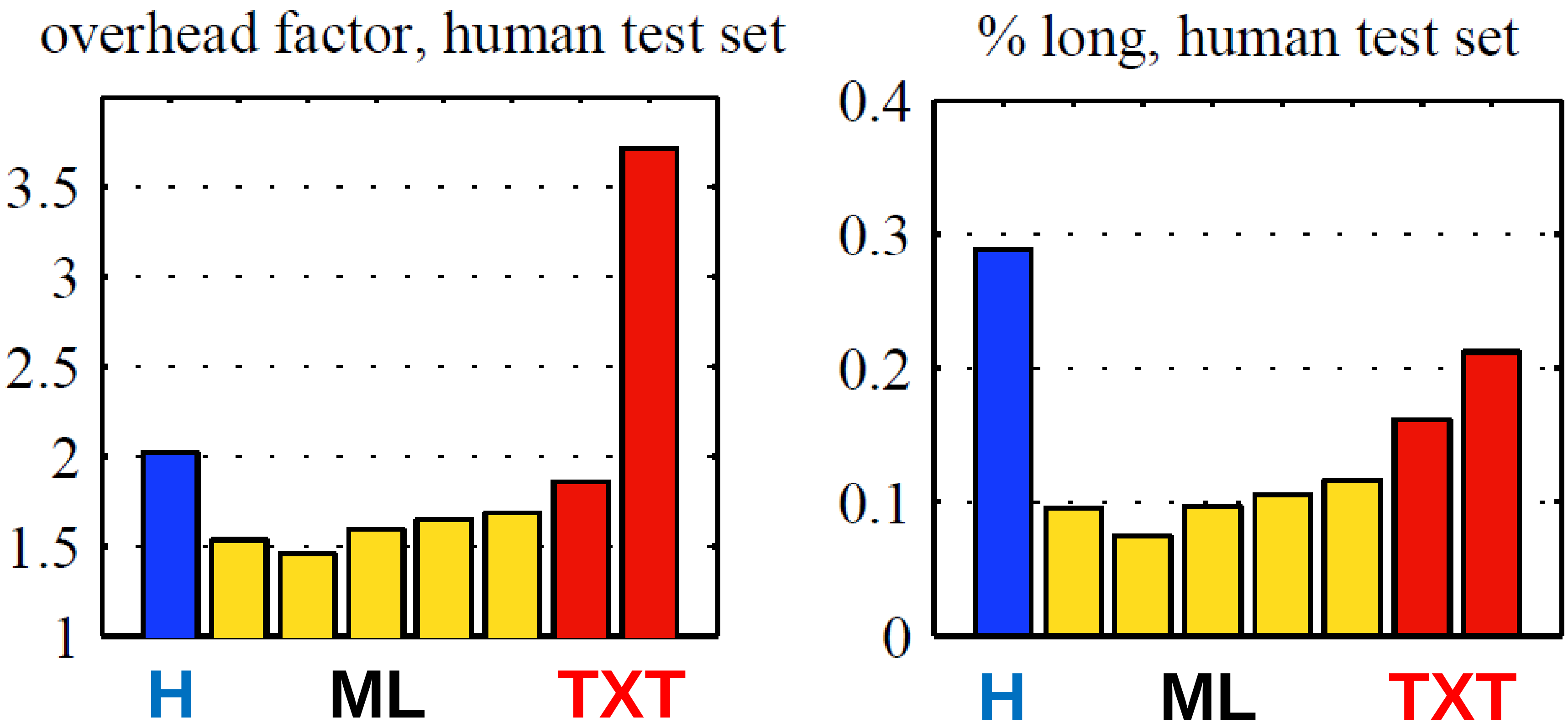
Machine Agents

- (1) Human
- (2) Similarity based (TXT):
  - $V(W|U,T) = \text{tf-idf}(W,T)$
    - Go to W that is textually most similar to T
- (3) Machine learning agents (ML):
    - Use human/shortest paths to learn the value function
  - Support Vector Machines
  - Reinforcement Learning

Machine Learning Agents

- **Features for the machine learning agents**
  Inspired by analysis of human behavior
  - sim(next, target) (TF-IDF cosine)
  - sim(current, next)
  - deg(next)
  - taxdist(next, target) (taxonomical distance)
  - linkcos(next, target) (cosine similarity in outgoing hyperlinks)

Results

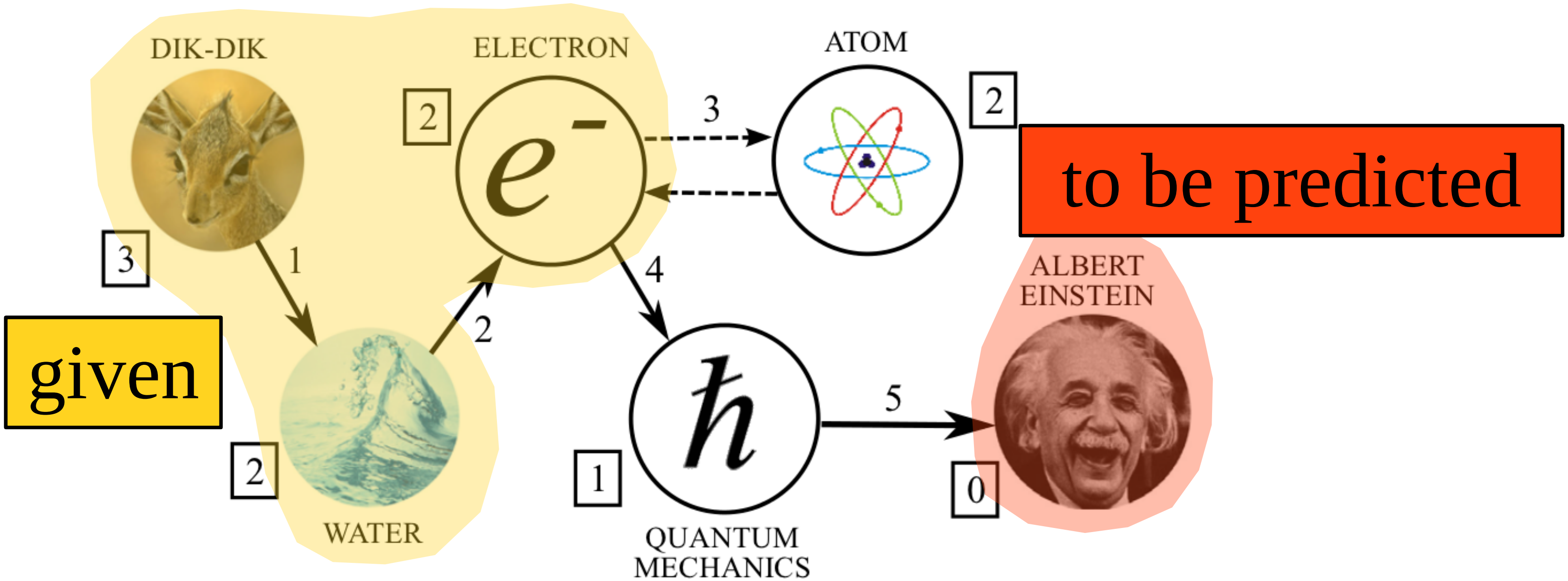


- **Machine** beats **human**!
- But, machines can get terribly lost
- Humans are sloppy (83% they miss a direct link)

Application: Target Prediction

Can we predict where the user is going?

Application: Target Prediction



- **Task:** Given first few clicks
- Predict the **target** player is trying to reach

Model of Navigation

- **Markov model of human navigation**

$$P(u_{i+1} | u_i, t; \Theta) \propto exp(\theta_i^\top \mathbf{f}(u_i, u_{i+1}, t))$$

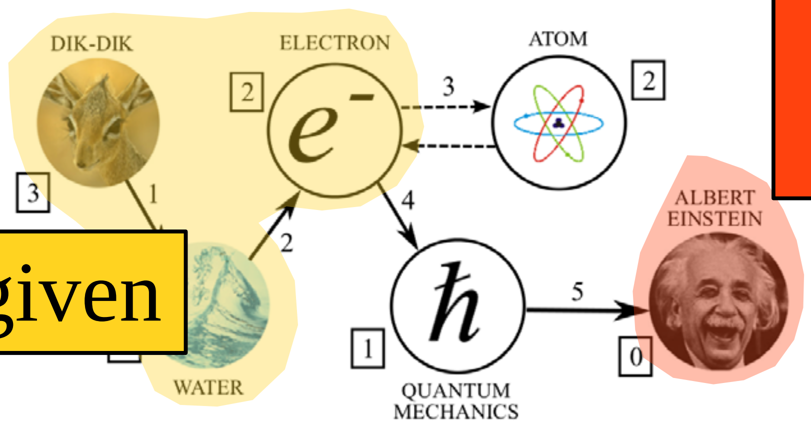
next — current — target — params — features

- **Predict the most likely target**

$$\arg\max_t P(q | t; \Theta) = \arg\max_t P(u_1) \prod_{i=1}^{k-1} P(u_{i+1} | u_i, t; \Theta)$$

given path prefix

Target Prediction: Training



given

to be predicted & given for training

- Fit Θ in learning-to-rank setup [Weston et al. '10]

initial Θ

Kimchi
Gopher
…
Albert Einstein
…

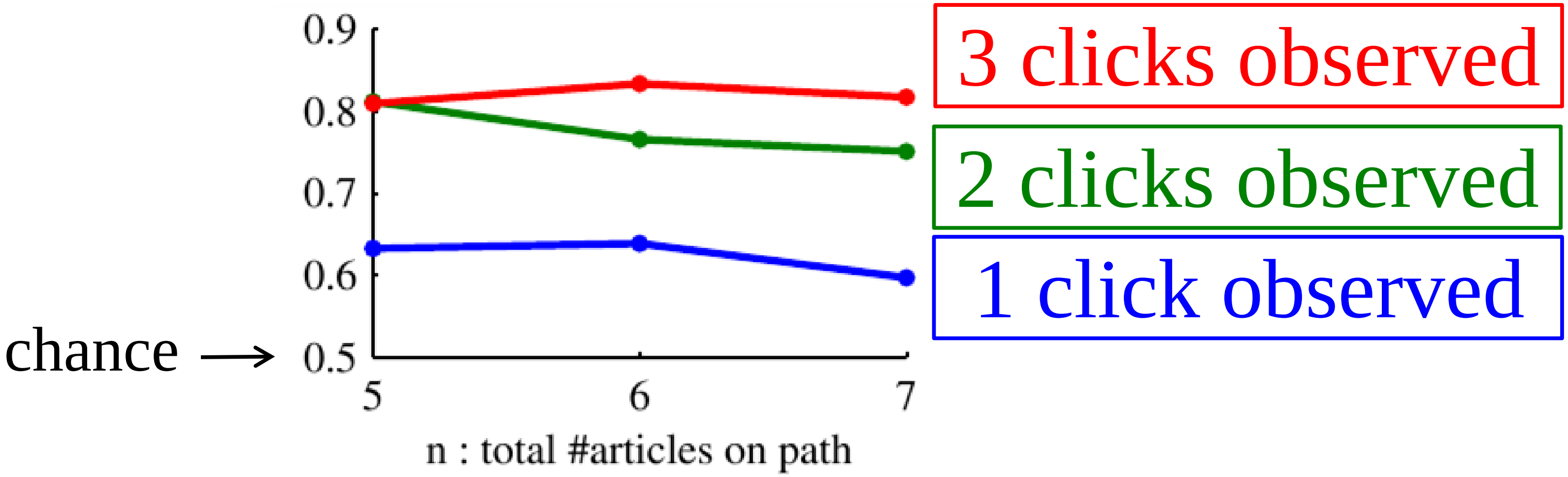
training

final Θ

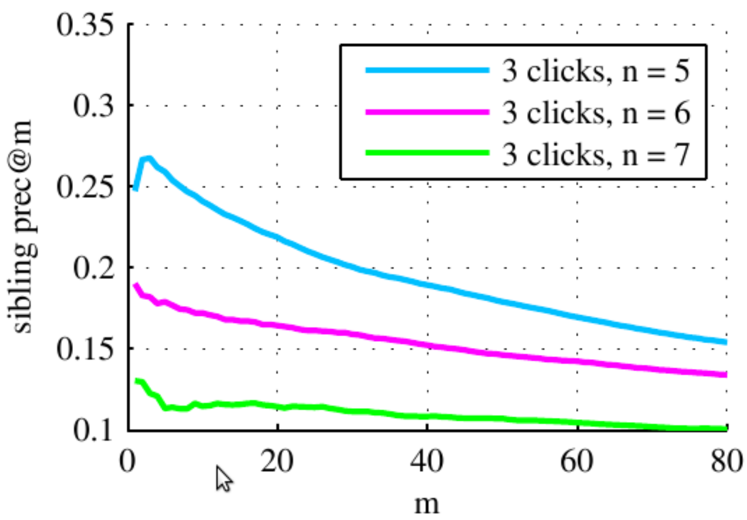
Albert Einstein
…
Football
Orpheus
…

Target Prediction: the Model

- Given choice of 2, choose true target



- Rank articles such that true target gets high rank



Conclusion: Navigation

- **Humans manage to find their ways in large networks, despite having only local information**
- **How do they do it?**
- Analyze large-scale data from the MSN network and Wikispeedia game
- **Answer:** They leverage expectations about network connectivity, based on background knowledge

Reflections: Navigation

- **Computational ideas play 2 crucial roles**
  - **Designing systems** in this new space
  - **Modeling** the social processes
- **Designing systems: Search engines**
  - User click-trails for web search ranking [Bilenko-White, '08]
  - Web revisitation patterns for crawling [Adar et al. '08]

Reflections: Navigation

- **Designing systems: Navigational tools**
  Is user lost? Where is she trying to go?
  - User facing tools and browsers: ScentTrails [Olston-Chi, '03]

  - Creating navigable networks
  - Navigable maps, ontologies [Helic-Strohmaier et al., '11]
  - Social browsing

Final Reflections

- **Models: How we search for information**
  - Information scent [Chi et al., '01]
  - Information foraging [Pirolli, '99]

- **Networks facilitate new ways of interacting with information**
  - Targeted search vs. Casual browsing
- **Can all this help us understand ourselves and each other any better?**


WHERE
are
YOU
going?
ameenafalchetto.com

THANKS!

Data + Code:

http://snap.stanford.edu